THOMAS NAGEL

Otras mentes

Ensayos críticos 1969-1994

BEG

BIBLIOTECA ECONÓMICA GEDISA

FILOSOFÍA

LECTURA Y CULTURA SIN EGOISMOS
BIBLIOTECA
www.lecturasinegoismo.com

Thomas Nagel

OTRAS MENTES

Serie: CLA•DE•MA
FILOSOFÍA

Editorial Gedisa ofrece
los siguientes títulos sobre

FILOSOFÍA Y CIENCIAS COGNITIVAS

THOMAS NAGEL	*Otras mentes* **Ensayos críticos 1969-1994**
DAVID J. CHALMERS	*La mente consciente* **En busca de una teoría fundamental**
JERRY A. FODOR	*Conceptos*
ERNST TUGENDHAT	*Diálogo en Leticia*
ERNST TUGENDHAT	*Ser-Verdad-Acción* **Ensayos filosóficos**
PAUL M. CHURCHLAND	*Materia y conciencia*
JAAKO HINTIKKA	*El viaje filosófico más largo* **De Aristóteles a Virginia Woolf**
D. C. DENNETT	*Contenido y conciencia*
IAN HACKING	*El surgimiento de la probabilidad*
JON ELSTER	*Lógica y sociedad*
HILARY PUTNAM	*Representación y realidad*
THOMAS NAGEL	*La última palabra*
KITARO NISHIDA	*Indagación del bien*
MARTÍN HEIDEGGER	*Introducción a la metafísica*
MANUEL CRUZ (COMP.)	*Tolerancia o barbarie*
IAN HACKING	*El surgimiento de la probabilidad*

OTRAS MENTES

Ensayos críticos
1969-1994

Thomas Nagel

gedisa
editorial

Título del original en inglés: *Other Minds. Critical Essays 1969-1994*
Publicado por Oxford University Press, New York, Oxford
© 1995 by Oxford University Press, Inc.

La presente traducción de *Otras mentes*, obra publicada originalmente en inglés, en 1995, se publica por acuerdo con Oxford University Press Inc.

Traducción: Sandra Girón

Diseño de cubierta: Juan Santana

Primera edición: octubre del 2000, Barcelona

CREATIVE COMMONS

Derechos para todas las ediciones en castellano

© Editorial Gedisa, 2000
Paseo Bonanova, 9 1º 1ª
08022 Barcelona, España
Tel. 93 253 09 04
Fax 93 253 09 05
correo electrónico: gedisa@gedisa.com
http://www.gedisa.com

ISBN: 84-7432-677-X
Depósito legal: B. 38085/2000

Impreso por: Carvigraf
Clot, 31 - Ripollet

Impreso en España
Printed in Spain

A la memoria de
Carlos Nino
1943-1993

Índice

Introducción: La cultura filosófica 11

I. FILOSOFÍA DE LA MENTE

1. El antropomorfismo de Freud 23
2. La revolución permanente de Freud 39
3. Wittgenstein: El conflicto egocéntrico 63
4. Chomsky: Lingüística y teoría del conocimiento 76
5. Fodor: Los límites del espacio interno 87
6. Armstrong: Sobre la mente 95
7. Dennett: Contenido y conciencia 107
8. Dennett: La disolución de la conciencia 112
9. O'Shaughnessy: La voluntad 117
10. Searle: ¿Por qué no somos ordenadores? 125

II. ÉTICA Y FILOSOFÍA POLÍTICA

11. Aristóteles: Sobre la *Eudaimonía* 145
12. Rawls: Sobre la justicia 154
13. Nozick: Libertarismo sin fundamentos 173
14. Hare: El pensamiento moral 189
15. Hare: Los fundamentos de la imparcialidad 196
16. Williams: Un pensamiento de más 210
17. Williams: Resistiendo a la teoría ética 218
18. Schelling: El precio de la vida 229
19. Schelling: Identidad personal y autodominio 235

20. Dworkin: Interpretación y derecho 242
21. MacIntyre contra la Ilustración 254
22. Kolakowski: La modernidad y el diablo 263

ÍNDICE TEMÁTICO 269

Introducción: La cultura filosófica

Estos ensayos fueron escritos a pedido, por lo que no tratan solamente acerca de los filósofos que más me interesan. Sin embargo, los temas son los que me han ocupado siempre: subjetividad y conciencia, objetividad y ética, liberalismo y razón. Aun cuando no se encuentre mucho sobre algunas de las partes que componen el núcleo principal de la filosofía —lenguaje, lógica, metafísica general— todo pertenece al universo común del argumento que define a la filosofía analítica: estos temas y los métodos usados para tratarlos no pueden aislarse unos de otros. He revisado levemente la mayoría de los ensayos para incluirlos en este libro y, en algunos casos, he introducido nuevas ideas, especialmente en los escritos más tempranamente. Pero he incluido sólo aquellas que todavía considero mayormente correctas, incluso después de décadas. Mis simpatías y antipatías básicas —antirreduccionistas y más o menos realistas— no han cambiado mucho.

Deseo agregar aquí algunas opiniones y reminiscencias personales sobre la cultura filosófica a la que pertenecen estos ensayos y casi todos sus temas, con el firme reconocimiento de que la sociología de un tema no reemplaza la reflexión sobre su sustancia. Cuando me inicié en ese campo, la filosofía analítica comenzaba apenas a insinuar su predominio en los Estados Unidos. La rama continental del análisis filosófico había ingresado con ayuda de emigrados como Carnap, Tarski, Reichenbach, Hempel y Feigl. La rama inglesa comenzaba a ejercer influencia, y la recibí porque asistí a Cornell como estudiante en 1954, cuando era probablemente el mejor lugar para aprender sobre Wittgenstein y la filosofía del lenguaje ordinario. Yo quería entonces graduarme en Física —el gran campo

romántico en esa época—, pero tomé un curso introductorio de Filosofía y pronto llegué a la conclusión de que mi verdadero motivo para estudiar física (a saber, descubrir si había tal cosa como la realidad objetiva), sin mencionar mis aptitudes, haría del departamento de Filosofía una elección más acertada.

En Cornell conocí a Norman Malcolm, John Rawls y Rogers Albritton, maestros que han ejercido sobre mí una fuerte y permanente influencia. Malcolm era un hombre robusto, encantador y vigoroso, cuya capacidad para mostrar su placer o desagrado sobre lo que se decía en la discusión filosófica era causa de gran ansiedad entre sus alumnos. Aunque su tendencia a moralizar el tema no era demasiado saludable, recibí de él y de Albritton la capacidad de experimentar el carácter distintivo y desconcertante de los problemas filosóficos y la importancia de Wittgenstein y de Rawls, la convicción de la prioridad de la teoría moral sustancial sobre la metaética. Mediante una beca Fulbright llegué a Oxford, que en 1958 estaba casi en la cima de su confianza y actividad como vanguardia de un movimiento. J. L. Austin fue mi supervisor de tesis para el grado de B.Phil.; era un catedrático brillante y cautivante, un crítico sarcástico y demoledor y un consejero maravillosamente atento, pero en realidad no estaba interesado en la clase de cuestiones que me atraían: desconfiaba de la "profundidad" y del estilo atormentado wittgensteniano, y la profundidad era lo que yo ansiaba. Austin falleció hacia el fin de mi segundo año allí, a la edad que ahora me parece increíble de cuarenta y ocho años. Dejando de lado su teoría de los actos de habla, su forma de estudio empírico detallado del lenguaje ordinario, aunque haya sido agradable de presenciar, no ha dejado muchas huellas, ni tampoco la convicción, entonces común, de que muchos de los problemas tradicionales de la filosofía podrían desaparecer aclarando confusiones sobre cómo funciona el lenguaje ordinario.

Paul Grice fue mi tutor durante un cuatrimestre. Fue la experiencia educativa más transformadora de mi vida. Yo iba una vez por semana con un artículo para leer en voz alta. Grice estaba sentado en su silla frente a mí, con aspecto aburrido y poco receptivo. Luego de leer algunas oraciones, él me hacía alguna pregunta breve que yo no podía responder. Finalmente, yo intentaba alguna respuesta, pero esta a su vez era puesta en duda, y pronto me veía reducido al silencio. Al final de una hora pasada casi totalmente en silencio, yo me retiraba

con una visión totalmente transformada, cualquiera que hubiese sido mi concepción anterior. El lograba mostrar, sin decir casi nada, que este o aquel enfoque corriente sobre algún problema tradicional no iba a funcionar. Nunca más tuve una experiencia similar.

Debido a su gran tamaño, con estudiantes de todo el mundo y muchos más seminarios de los que uno podía tomar, y debido a su gran empeño tradicional en enseñar a través de la discusión, Oxford era una fiesta. Pasaban muchas cosas: las clases de Strawson, que se convirtieron en *Individuals;* las de Hart, que se convirtieron en *The Concept of Law;* las de Anscombe sobre Wittgenstein y la teoría de la acción, las de Philippa Foot sobre creencia moral y motivación. Asistí a valiosos cursos tutoriales con David Pears, James Thomson y G.E.L. Owen. Blackwell's era la mejor librería de filosofía del mundo, y los libros eran tan baratos como lo son hoy las calculadoras de bolsillo. Era un lugar maravilloso para estar en el momento de empezar a abrirse camino y tratar de decidirlo todo.

Al cabo de dos años regresé a los Estados Unidos, a la menos vivificante Harvard. Quine, descendiendo de la tradición del empirismo lógico reconstructivo (la otra gran rama de la dinastía de Frege), estaba ocupado con un proyecto filosófico, deslumbrante para mí, de volver a describir el mundo. Escribí mi tesis con Rawls (quien se había trasladado a Harvard) y tuve la oportunidad de leer y discutir con él sucesivos borradores de *A Theory of Justice.* Conocí a Noam Chomsky y Saul Kripke, que era entonces estudiante, pero ya distinguido por su trabajo en lógica. Gilbert Harman, Robert Swartz (estudiantes de posgrado) y yo nos encontrábamos regularmente para discutir las *Investigaciones filosóficas* de Wittgenstein. La nueva interpretación que Kripke estaba comenzando a elaborar modificó mi comprensión del libro.

Harman y yo caminábamos por la avenida Massachusetts hacia el MIT varias veces por semana para escuchar las clases de Chomsky. (El y Austin tenían el notable poder de lograr que uno estuviera pendiente de cada una de sus palabras, sin la menor actuación.) Se trataba de las clases que luego constituyeron *Cartesian Linguistics* y *Aspects of the Theory of Syntax.* Aunque me exasperaban a veces sus polémicas con otros escritores, la obra de Chomsky tuvo en mí gran impacto. Nunca aprendí mucho de lingüística. Lo que me atraía era su extraor-

dinaria imaginación y su actitud hacia la mente como un campo abierto a la investigación teórica y empírica combinadas, así como su fuerte impresión de que estábamos sólo al comienzo de nuestra comprensión y de que podríamos ser por constitución incapaces de hacer en este campo el tipo de progreso que los humanos pudieron hacer en física. También me interesaba la falta de "pureza" de su enfoque. A él no le preocupaban los límites entre las cuestiones y métodos empíricos, teóricos y filosóficos. Si bien estas distinciones son a menudo importantes, a él parecía interesarle todo al mismo tiempo.

En 1963 me trasladé a Berkeley, con mi primer empleo. Allí fui influido especialmente por la técnica de Thompson Clarke de convertir la naturaleza de los problemas filosóficos en sí mismos en un objeto de investigación y un recurso para la comprensión de los conceptos humanos básicos y de las formas de pensamiento. La década de 1960 comenzó con brío en Berkeley poco después de que yo llegara allí y, luego de tres años, regresé al este, a Princeton, que estaba en proceso de convertirse en un departamento grande y poderoso. Investigué activamente en varias áreas, incluyendo ética e historia de la filosofía, pero el centro de gravedad era una mezcla de metafísica, epistemología, filosofía del lenguaje, filosofía de la lógica y matemática y filosofía de la ciencia. En los últimos años se enfocan estos temas desde un espíritu de construcción de teorías que concibe a la filosofía como continuación de la ciencia, sólo que más abstracta y general. Esta perspectiva, que nunca me atrajo, es ahora muy común entre los filósofos analíticos, incluyendo a muchas de las mejores mentes sobre el tema; de algún modo la tradición Carnap-Quine ha llegado a dominar la profesión, aunque la otra, que se centra en los problemas, existe todavía y creo que es importante mantenerla viva. Sin la fuerte comprensión del carácter originalmente filosófico de ciertos problemas, es demasiado fácil quedar atrapados en el cientificismo: la idea de que cualquier cuestión genuina puede ser tratada como parte del desarrollo de una cosmovisión científica, y de que aquella que no lo puede no es una cuestión real. Esta perspectiva puede llevar a trabajos de sorprendente superficialidad, como el materialismo eliminativo en la filosofía de la mente.

El restablecimiento de la teoría política sustancial, por parte de Rawls, y el ascenso de la temperatura política asociado a

la Guerra de Vietnam, el movimiento de los derechos civiles y la discusión sobre el aborto, afectaron el desarrollo de los trabajos en otros ámbitos de la disciplina. A finales de la década de 1960, Marshall Cohen, T. M. Scanlon y yo iniciamos *Philosophy & Public Affairs* (aunque el primer número no apareció hasta 1971); además Robert Nozick y yo, alentados por Rawls, iniciamos un grupo de discusión mensual llamado SELF (The Society for Ethical and Legal Philosophy). La mayoría de los miembros éramos ya amigos, pero había además otros, como Ronald Dworkin, Michael Walzer, Charles Fried, Judith Thomson y Frank Michelman, a quienes pude conocer bien gracias a estas reuniones. La rapidez y el nivel de discusión eran excepcionales, y varias obras significativas en curso pasaron por este grupo, unido por la confianza de que se puede combinar la intuición moral con el razonamiento sistemático para lograr un progreso sustancial en temas normativos reales.

En suma, yo estaba muy feliz con la camaradería filosófica que encontré en los primeros años. A los veinte y a comienzos de mis treinta años pasé algunos de los mejores momentos de mi vida discutiendo filosofía con amigos: Tom Clarke, John Searle, Barry Stroud, en Berkeley; Gil Harman y Tim Scanlon, en Princeton; Saul Kripke y Bob Nozick, en Nueva York (cuando ellos estaban en la Universidad Rockefeller y yo vivía cerca, en Manhattan); y Ronald Dworkin, Derek Parfit, David Wiggins y Bernard Williams en ambos lados del Atlántico. A esa edad uno tiene energía para consumir, no economiza todavía los recursos, y todo está por decidir. Existe una clase de discusión muy regocijante, a menudo muy rápida, que se da sólo entre personas en la misma sintonía y que conocen el tema exhaustivamente. Esto se vuelve más raro a medida que se envejece —las sinapsis se cansan, las categorías se endurecen y la capacidad de entusiasmo se debilita—, pero desempeña un papel importante para la mayoría de nosotros en el desarrollo y difusión de nuestras ideas filosóficas.

Es por eso que la filosofía de este tipo es a menudo un fenómeno local, a pesar de la universalidad del interés. Así como los pintores que miran la obra de otros, los filósofos analíticos hablan entre sí, y los desarrollos originales han convocado a pequeños grupos en una ciudad o en lugares de fácil acceso. Posteriormente, a través de becas de viaje y por emigración, la

tradición se difundió desde Viena, Cambridge y Oxford, primero a todos los países angloparlantes y Escandinavia, y luego a todo el mundo. Ha sido un placer participar en este flujo de ideas y hacerme de amigos en muchos lugares. Dedico este libro a la memoria de uno de ellos, a quien aprecié mucho —un hombre de incomparable grandeza de corazón, dulzura de carácter, valentía, alegría e inteligencia—, Carlos Nino, de Buenos Aires, quien murió en 1993 a la edad de cuarenta y nueve años, luego de haber desempeñado un papel esencial en la vida intelectual y pública de su país y en conducirlo hacia el estado de derecho.

Una vez me entrevistaron dos filósofos franceses, que viajaban por los Estados Unidos haciendo una encuesta sobre la disciplina. Cuando me preguntaron acerca de mis primeros intereses y les respondí que había pasado de la física a la filosofía, se sonrieron intencionadamente. Cuando les pregunté por qué, me contestaron que todos los filósofos estadounidenses que habían entrevistado les habían respondido lo mismo: su primer interés había sido la ciencia o la matemática. En mi caso, esto se debía en parte a que esos eran los únicos temas que ofrecían estímulos intelectuales en la escuela a la que yo había asistido: no se los podía arruinar con ignorancia, incompetencia ni propaganda política. Pero también me encantaba el rigor y la universalidad de las ciencias físicas y la matemática, y nunca me abandonó esta actitud. En tanto que un amor temprano hacia la ciencia no caracteriza a todos los filósofos analíticos norteamericanos, lo es para muchos, y tiene consecuencias en la disciplina. Los entusiastas de la ciencia adolescentes son un tipo intelectual diferente de aquellos con intereses históricos o literarios, que es el trasfondo más habitual en Europa. En los Estados Unidos, los filósofos no tienen que saber demasiado sobre historia o literatura, pero sí se espera que sepan algo de ciencia, que tengan al menos la comprensión del aficionado sobre las contribuciones de Newton, Maxwell, Darwin, Einstein, Heisenberg, Cantor, Gödel y Turing, así como de las de los representantes de las ciencias humanas como Chomsky, Sperry y Arrow, quienes proporcionan datos para la reflexión filosófica. Es también importante el concepto de la filosofía no como un tema crítico o histórico, sino como una empresa de investigación primaria en la que la mayoría de nosotros estamos involucrados, aunque sea de una manera ínfima. Un *ethos*

más o menos científico significa que todo debe ser explícitamente expuesto y, de ser posible, discutido, y que la fuerza persuasiva del argumento es lo que importa; no se puede permitir que la referencia retórica y textual conduzca a la inercia. Este es al menos el ideal, aunque a veces se lo viole.

La comunidad filosófica que me formó, con todas sus variedades de intereses y perspectivas, también valoraba al argumento filosófico como una actividad, en algunos casos más que la filosofía como producto. Esto era paralelo a una apreciación inusual de la calidad activa de la mente de la gente, cualquiera que fueran sus opiniones o su cantidad de publicaciones. La filosofía analítica es maravillosamente hospitalaria —en mi opinión, más que cualquier otro campo académico— hacia individuos con obvio talento y que han publicado algunos artículos más o menos distinguidos, y también hacia filósofos que adquieren velocidad creativa gradualmente, que publican poco al comienzo de su carrera y, luego, a una edad mediana aparecen con obras importantes (por ejemplo, Rawls, Grice y Donald Davidson).

Es también una disciplina en la que las mentes agudas —C.I., velocidad lógica, agilidad mental— desempeñan un papel fundamental, aun cuando dichas aptitudes no sean lo mismo que la habilidad filosófica. La filosofía es como el baloncesto: ser alto no garantiza ser un buen jugador de baloncesto, pero ayuda mucho. En filosofía ayuda ser muy sagaz. Las mentes sagaces pueden viajar más lejos y más rápido que el resto de nosotros, y sería deseable que la filosofía atrajera a mayor cantidad de ellas. Pero los efectos de este tipo de inteligencia son complejos: a veces, si todo ese poder está al servicio de intuiciones atolondradas, lleva a resultados lógicamente deslumbrantes pero inaceptables. Sin embargo, aun cuando se salga de los rieles, la brillantez genera estructuras de pensamiento que requieren atención y tienen vida propia, y su impacto en la disciplina no depende de que alguien las considere correctas. Esto puede ser una nimiedad, pero supongo que la desvalorización de su posibilidad es inevitable en una disciplina tan dominada por el argumento.

Un determinante crucial del carácter de la filosofía analítica —que es una suerte, en lo que a mí respecta— es la poca importancia, en el mundo angloparlante, del intelectual como figura pública. La fama no importa, y dar una opinión sobre

casi todo no es parte del trabajo. No es necesario que los filósofos pertenezcan "más a su época" de lo que deseen; no tienen que escribir para el mundo, pero pueden buscar cuestiones dentro de su tema, con el grado de dificultad que estas exijan. Si el trabajo es de alta calidad, recibirá el apoyo de una academia grande y dedicada, que es generalmente independiente de la opinión popular. Es una posición envidiable, en comparación con escritores cuyo lugar e ingresos dependen de la reacción de un público más amplio. Claro que hay modas superficiales y puntos ciegos dentro de la comunidad académica, pero en filosofía, al menos, su efecto no es tan negativo como sería la necesidad de competir por la fama literaria. Creo que son preferibles los tecnicismos más áridos a la mezcla de sobresimplificación y falsa profundidad, que es una forma demasiado a menudo adoptada por la filosofía popular. Una academia fuerte provee un valioso refugio para la tarea difícil y a menudo muy especializada que se debe hacer para poder adelantar en el tema.

Ultimamente, algunos sostenedores de la "filosofía fácil" se han vuelto mundialmente famosos, y ha habido un surgimiento del filosofar aficionado fuera de la disciplina, que no la ha afectado por dentro. La filosofía analítica ha podido escapar casi completamente del relativismo fácil que parece tener tanta influencia en las otras ciencias humanas, originalmente iniciado por Derrida y ahora defendido con referencias a Richard Rorty, Paul Feyerabend y Thomas Kuhn. La filosofía parece exportar sus peores productos, como ocurrió a comienzos del siglo con el conductismo y el positivismo lógico, y con su influencia en las ciencias sociales y la psicología.

Tendencias como esta plantean una elección difícil. Cuando una filosofía denigrada ejerce gran influencia en otras partes, la única manera de combatirla activamente es entrar en la arena y competir para lograr convencimiento popular, en otras palabras, asumir el papel tradicional del intelectual público. Si bien admiro a aquellos como Dworkin y Searle, quienes tienen estómago y talento para este tipo de polémica, yo perdí el deseo —si es que alguna vez lo he tenido— y espero que esta ola de confusión se aquiete si la ignoramos. Creo que es un gran peso para los profesionales tener que competir para lograr la aprobación de quienes no entienden el tema. A veces es inevitable hacerlo, como cuando se necesita apoyo del Estado

para mantener una investigación científica, pero por suerte la filosofía no cuesta demasiado.

Ciertamente, ha habido conexiones útiles entre la Filosofía y otras disciplinas: Derecho, Economía, Política, Medicina, Lingüística, Psicología y Tecnología informática, pero que se deben al interés común en temas específicos, más que a la exportación de doctrinas epistemológicas o metafísicas para todo uso. Y siempre que estas conexiones sean importantes y duraderas, la vida real de la disciplina depende del trabajo llevado a cabo en relación con sus problemas centrales.

Sustancialmente, ha habido excelentes buenos trabajos, pero, ¿cuán buenos? ¿Cuáles escritos filosóficos de este siglo tendrán interés más que histórico dentro de cien años? Cerca del final del siglo XX se leen muy pocos filósofos del siglo XIX: Peirce y Frege, Mill, Bentham, Sidgwick, Hegel, Nietzsche y quizá Schopenhauer. ¿Cómo sería una lista comparable la próxima vez? Pero, a su vez, ¿quién hubiera apostado a Peirce, Frege o Nietzsche en 1894? La grandeza es rara: el siglo IV a.C. y los siglos XVII y XVIII d.C. no han sido superados, en mi opinión. El más fuerte candidato reciente para la inmortalidad es Wittgenstein, quien ciertamente identificó una nueva serie de problemas, aunque sus respuestas a ellos todavía no son bien comprendidas y muy difíciles de evaluar. Russell, Husserl, Sartre y Carnap serán nombres familiares por un tiempo, así como —lamentablemente— el de Heidegger. Sin embargo, entre los que florecieron en la segunda mitad de este siglo, es difícil adivinar quién va a ser realmente *leído* de aquí a un siglo. Apostaría a Rawls, y quizá Nozick, debido en parte al carácter lineal de la historia del pensamiento político occidental y a su necesidad de figuras ejemplares, pero no confío en que los trabajos más influyentes en metafísica, epistemología y filosofía del lenguaje sobrevivan. Pero tal vez se trata de una pregunta equivocada: tal vez sea posible contribuir en gran medida al desarrollo de la filosofía sin escribir un clásico.

Una vez Bernard Williams hizo una pregunta embarazosa: ¿Cuál es la razón de hacer filosofía si uno no es extraordinariamente bueno para ello? El problema es que, a diferencia de un historiador modestamente dotado, uno no puede hacer sólidos descubrimientos, sobre los que otros puedan apoyarse para lograr resultados más importantes. Si no somos extraordinarios, lo que hacemos en filosofía será poco original (y, por lo tanto,

innecesario) o inadecuadamente fundamentado (y, por lo tanto, inútil). Más probablemente, será poco original y falso. Esta es la razón por la que no vale la pena estudiar la mayor parte de los trabajos filosóficos del pasado. Entonces, ¿no es absurdo pagar a miles de personas para que piensen acerca de estas cuestiones fundamentales?

Si existe una respuesta a esta pregunta, debería depender de la idea de que estamos embarcados en una empresa colectiva, cuyos resultados no pueden ser siempre fácilmente rastreados. Alguna clase de intercambio de argumentos e ideas puede generar desarrollos de valor que no podrían llevarse a cabo aun con los más grandes pensadores trabajando individualmente y respondiéndose mutuamente. De este modo, se divulgan e interpretan las contribuciones originales y se genera un medio ambiente en el cual otros pueden desarrollarse ocasional e impredeciblemente. No estoy seguro de que esto sea cierto, pero sí me parece posible que las actividades cooperativas de la investigación y docencia filosóficas, ampliamente extendidas, y gracias a las cuales esta disciplina está viva, contribuyen en algo al desarrollo de las herramientas intelectuales que, en el largo plazo, mejoran nuestra capacidad común para pensar sobre el lenguaje, la ciencia, el conocimiento, la política, la moral, y acerca de lo que es el ser humano.

De todos modos, algo sucederá próximamente. Siempre es así.

I

FILOSOFÍA DE LA MENTE

1

El antropomorfismo de Freud*

Es difícil decidir si una psicología mentalista, que trata procesos tanto conscientes como inconscientes, puede ser considerada una teoría de orden superior sobre el funcionamiento del cerebro. La psicología se ocupa de personas y de las causas e influencias que explican lo que ellas hacen, piensan, sienten, etc. Estas causas e influencias deben tener alguna base fisiológica, pero no está claro cuál debe ser la relación entre estos niveles de explicación, a fin de que ambos sean verdaderos. Sospecho que una teoría psicológica no necesita ser un tosco esbozo de la estructura física o fisiológica subyacente. En tanto no haya ninguna contradicción en el nivel de las conexiones causales particulares, es posible que ambos órdenes de descripción y explicación puedan ser verdaderos, sin que haya nada parecido a una relación de reducción entre las teorías. Esto se aplica tanto a la psicología de sentido común u ordinaria, como a su extensión psicoanalítica. Que esta haya sido también la posición de Freud, no lo sé.

Freud era materialista, y en una etapa temprana de sus investigaciones psicológicas intentó construir una psicología explícitamente fisiológica basada en la interacción de neuronas. Este intento, que ahora es muy conocido bajo el título "Proyecto para una psicología científica", fue abandonado poco después de que Freud enviara el borrador a Fliess en octubre de 1895. Cuando se enteró en 1937 de que Marie Bonaparte había sacado a luz el manuscrito, trató de que fuera destruido.

* Publicado en *Freud: A Collection of Critical Essays,* Richard Wollheim, comp. (Doubleday, 1974).

Sus teorías posteriores fueron de un carácter muy diferente, puesto que contenían sólo terminología psicológica y no se referían explícitamente a la interacción de neuronas. Sin embargo, hay una gran continuidad estructural entre las primeras visiones de Freud y las posteriores, y él siguió estando convencido de que el aparato psíquico que investigaba y describía en términos mentalistas era en su verdadera naturaleza un sistema físico, aunque se sabía demasiado poco de neurofisiología como para que alguien pudiera pensar en psicología en términos físicos. Por ello Freud consideró necesario abandonar la línea de investigación comenzada en el "Proyecto".

Por consiguiente, surge la pregunta de en qué sentido es posible reflexionar sobre un sistema físico en términos mentalistas, extraídos del vocabulario de experiencia, percepción, deseo y otros similares, sin tener idea del significado físico de esas descripciones. Esta pregunta atañe no sólo a la teoría psicoanalítica, sino también a las disputas actuales sobre el estado de hipótesis mentalistas en lingüística y en otras áreas donde se sostiene que es posible suponer una realización empírica de un proceso o función descrito mentalmente o antropomórficamente. ¿Cuál es el significado de tales afirmaciones?

Freud no se mantuvo callado en este tema, y sus explicaciones de cómo es posible pensar antropomórficamente un sistema físico cuando se carece de una comprensión explícitamente física de dicho sistema, se hallan entre los pasajes más filosóficos de sus escritos.[1] En ellos se encuentra también una contribución a la discusión del problema mente-cuerpo, que vale la pena examinar.

Las notas de las que nos ocuparemos forman parte de la defensa general de Freud de la existencia de los estados mentales inconscientes. Se ubican principalmente en los siguientes cuatro escritos: *La interpretación de los sueños,*[2] "El inconsciente",[3] *Esquema del psicoanálisis,*[4] y "Algunas lecciones elementales de psicoanálisis".[5] Será útil citar un pasaje típico completo.

> Nuestra noción de un aparato psíquico de dimensión espacial, adecuadamente integrado y desarrollado bajo el influjo de las necesidades vitales; un aparato que sólo en un determinado punto y bajo

ciertas condiciones da origen a los fenómenos de conciencia, nos ha permitido estructurar la psicología sobre una base semejante a la de cualquier otra ciencia natural como, por ejemplo, la física. Esta, como aquella, persiguen el fin de revelar tras las propiedades (cualidades) del objeto investigado, que se dan directamente a nuestra percepción, algo que sea más independiente de la receptividad selectiva de nuestros órganos sensoriales y que se aproxima más al supuesto estado real de cosas. No esperemos captar este último, pues, según vemos, toda nueva revelación psicológica debe volver a traducirse al lenguaje de nuestras percepciones, del cual, evidentemente, no podemos liberarnos. He aquí la esencia y la limitación de la psicología. Es como si en la física declarásemos: contando con la suficiente agudeza visual, comprobaríamos que un cuerpo, sólido al parecer, consiste de partículas de determinada forma, dimensión y situación. Entretanto, tratamos de llevar al máximo, mediante recursos artificiales, la capacidad de rendimiento de nuestros órganos sensoriales, pero cabe esperar que todos estos esfuerzos nada cambiarán en definitiva. Lo real siempre seguirá siendo "incognoscible".

La elaboración intelectual de nuestras percepciones sensoriales primarias nos permite reconocer en el mundo exterior relaciones y dependencias que pueden ser reproducidas o reflejadas fielmente en el mundo interior de nuestro pensamiento, poniéndonos su conocimiento en situación de "comprender" algo en el mundo exterior, de preverlo y, posiblemente, de modificarlo. Así procedemos también en psicoanálisis. Hemos hallado recursos técnicos que permiten colmar las lagunas en nuestros fenómenos conscientes, y los utilizamos tal como los físicos emplean el experimento. Por ese camino elucidamos una serie de procesos que, en sí mismos, son "incognoscibles"; los insertamos en la serie de los que nos son conscientes, y si afirmamos, por ejemplo, la intervención de un determinado recuerdo inconsciente, sólo queremos decir que ha sucedido algo absolutamente incaptable para nosotros, pero algo que, si hubiese llegado a nuestra conciencia, sólo habría podido ser así, y no de otro modo.[6]

Freud parece llegar a esta posición a través del siguiente proceso de razonamiento: tratar de construir una ciencia de la psicología que se ocupe sólo de procesos conscientes parece una tarea imposible, ya que hay demasiados resquicios causales evidentes. El material consciente es fragmentario y asistemático, y por lo tanto es improbable que pueda ser comprendido teóricamente en términos que no vayan más allá.

Es natural suponer que estos resquicios se llenan con procesos neurofisiológicos que producen, de vez en cuando, los estados conscientes. Y los propósitos de unidad teórica se satisfacen suponiendo, en vez de una alternancia e interacción entre procesos físicos inconscientes y procesos mentales conscientes, un sistema causalmente completo, en el cual, sin embargo, algunos procesos tienen además la propiedad de conciencia, o tienen concomitantes conscientes. Lo mental aparece, entonces, como el efecto de cierto tipo de proceso físico.[7]

Sin embargo, la reflexión posterior sugiere que puede ser un error identificarlo con los procesos físicos mismos. Estos pueden aparecer a la conciencia, pero son en sí mismos inconscientes, como todos los procesos fisiológicos. Y, dado que la verdadera naturaleza de los procesos mentales que aparecen a la conciencia es física, y siendo la conciencia simplemente una cualidad que se les agrega, no puede haber ninguna objeción a considerar mentales aquellos procesos intermedios, que ocurren en el mismo sistema físico, y que no aparecen a la conciencia aunque puedan ser en muchos aspectos funcionalmente similares a los que sí lo hacen. Más aún, ya que no tenemos la comprensión física del sistema nervioso necesaria para poder pensar acerca de estos procesos en términos físicos (tal vez nunca podremos reducirlos a términos *celulares)* nuestra mayor esperanza de progreso en la comprensión del sistema físico es pensarlo en términos de los aspectos conscientes bajo los cuales algunos procesos mentales se nos aparecen. Esto es análogo a nuestro uso de imágenes en la reflexión acerca de la física, aun sobre fenómenos físicos que no son de hecho visibles. El reflexionar sobre procesos mentales en términos de aparición a la conciencia no implica que su naturaleza intrínseca sea consciente. De hecho, la naturaleza intrínseca de los procesos mentales tanto conscientes como inconscientes nos es desconocida, y ambos son meramente representados, en forma no exhaustiva, por imágenes conscientes. Por lo tanto, todo lo psíquico, y no sólo lo inconsciente, es *en sí mismo* inconsciente.

Así como Kant nos advirtió no dejar de lado el hecho de que nuestras percepciones están condicionadas subjetivamente y no deben ser consideradas idénticas a lo que es percibido aunque imposible de conocer, así el psicoanálisis nos advierte no equiparar a las per-

cepciones por medio de la conciencia con los procesos mentales inconscientes que son su objeto. Análogamente a lo físico, lo psíquico no es necesariamente en realidad lo que nos parece ser.[8]

Deseo considerar tres cuestiones acerca de esta posición. Primero, ¿es adecuada la analogía con el empleo de imágenes en la física? Segundo, ¿implica esta concepción una posición particular sobre el problema mente-cuerpo (por ejemplo, el materialismo), o es compatible con diferentes alternativas? Tercero, ¿nos provee una razón para el uso de conceptos mentales, tomados de la psicología de la conciencia, para teorizar acerca de aquellos procesos, de cuya naturaleza fisiológica o química no podemos formarnos una concepción?

Es verdad, por cierto, que las imágenes visuales son útiles al reflexionar acerca de estructuras invisibles, invisibles porque son demasiado pequeñas o porque no reflejan la luz. Por eso podemos imaginarnos una molécula de ADN como una doble espiral. ¿Esto quiere decir que creemos que la veríamos de este modo si nuestra vista fuera lo suficientemente aguda? Tal vez sí. Pero, para algunos objetos, tales como los núcleos atómicos, suponer que nuestra vista pueda ser lo suficientemente aguda como para poder *ver* su estructura no tiene mucho sentido. Es más plausible suponer que *si* creemos la proposición hipotética, es porque creemos otra cosa: que hay similitud de estructura entre la cosa invisible de la que hablamos y otras cosas visibles que parecen ser de cierta manera, y que este rasgo estructural es responsable de que parezcan de ese modo. Si este rasgo estructural es lo que vemos en el caso de objetos visibles, entonces podemos emplear este tipo de imagen visual para representarnos el mismo rasgo estructural en objetos invisibles. De ahí nuestra imagen de la molécula de ADN.

Un aspecto importante de tales casos es que la estructura imaginada puede ser caracterizada en forma independiente. Una doble espiral puede ser descrita en términos puramente geométricos, sin referencia a su apariencia visual, y es lo primero, y no lo segundo, lo que tienen en común la molécula de ADN y un modelo visible.[9] Pero, si el significado del enunciado hipotético "si pudiéramos verlo, tendría este aspecto" depende de la disponibilidad de una caracterización independiente en términos no visuales, entonces la utilidad de este ejemplo, en tanto analogía con la relación entre los procesos conscientes e

inconscientes, es problemática. Ya que en este último caso no tenemos ninguna manera independiente de caracterizar el proceso mental inconsciente, "que, si hubiera ingresado en nuestra conciencia, podría sólo haber sido descrito de tal y tal manera".

Sin embargo, todavía podemos darle sentido a esta suposición en términos de la *posibilidad* de una caracterización independiente del proceso inconsciente. Podemos suponer que, aunque seamos por el momento totalmente incapaces de formarnos una concepción de él, de todos modos comparte rasgos con un proceso mental consciente que le corresponde, y estos rasgos son parcialmente responsables de que dicho proceso aparezca a la conciencia en la forma en que lo hace. Esta suposición parece legitimar el condicional contrafáctico peculiar, aunque no tengamos ahora el vocabulario o los conceptos para describir los rasgos comunes. Por ejemplo, no tienen que ser rasgos descritos en términos de la neurofisiología actual. Pueden ser descritos en términos de una psicología futura cuya forma esté determinada en parte por el desarrollo de la teoría psicoanalítica. Y puede ser que no haya reducción de los términos generales de esa teoría a los términos de la neurofisiología actual (aunque es posible que Freud mismo pensara que sí). Creo que esta interpretación hace comprensible el discurso mentalista de Freud, referido a lo que él considera un sistema físico inteligible; también hace aceptable la analogía con la visualización en física, aunque no tan cercana como pueda parecer al comienzo. En vez de inferir causas específicas similares a partir de efectos similares, él infiere *semejanza* de causas en aspectos desconocidos a partir de la semejanza de efectos observada.

Nuestro conocimiento del inconsciente, dice, es muy parecido a nuestro conocimiento de la mente de otros, ya que se apoya en fundamentos circunstanciales y de conducta.[10] Ya que tenemos razones estándar de este tipo para creer que ciertos rasgos de nuestra conducta y de la de otros tienen una explicación física, y otras razones para negar que estos fenómenos psíquicos son conscientes,[11] la conclusión natural es que son inconscientes, pero similares en otros aspectos a los procesos mentales potencialmente o actualmente conscientes a los que nuestras explicaciones habituales se refieren. Ya que no aparecen a la conciencia en forma similar, la semejanza debe ser hallada en otras características, probablemente físicas.

Ya que la conciencia no abarca la naturaleza de, por ejemplo, la hostilidad consciente, es posible emplear las imágenes de la conciencia para reflexionar sobre lo que es común tanto a las formas conscientes como a las inconscientes, cuando usamos imágenes visuales al pensar sobre la estructura de espirales dobles tanto visibles como invisibles.

La principal dificultad de esta posición es que puede suponer demasiado, aun cuando lo que supone es menos específico que en el caso de las imágenes visuales de entidades submicroscópicas o invisibles. Supone que *hay* algún carácter objetivo definido o una clase disyuntiva de caracteres comunes a los estados que habitualmente se agrupan por su semejanza en cuanto a su manifestación a la conciencia (y sus conexiones contextuales y de conducta y significado). Sólo si esto es verdad podemos remitirnos a un tipo de estado del sistema nervioso con un concepto mentalista, sin implicar nada respecto de sus manifestaciones conscientes. Pero esto no es evidente, de ninguna manera, al menos en muchos de los ejemplos importantes para Freud, como creencias, deseos, identificaciones.

Es de lo más improbable, por cierto, que haya un carácter general *neurológico* o una clase de caracteres comunes a todas las instancias del deseo de matar al padre, pero este no es el problema. Un defensor de la concepción freudiana no tiene que sostener que el carácter objetivo de estos estados puede ser explicado en términos de conceptos físicos existentes. Quizá sea sólo una psicología desarrollada, no reductible a la neurofisiología actual, la que puede dar cuenta de ellos.[12] Pero incluso suponer esto —es decir, suponer que una psicología objetiva, cuyos conceptos se refieren a fenómenos físicos, preserva aproximadamente las distinciones y categorías comprendidas en conceptos mentales de sentido común— es suponer mucho. (Es, tal vez, menos improbable en el caso de sensaciones que en el caso de estados mentales relacionados con el pensamiento [*thought-related*].) Si esta crítica es correcta, y las suposiciones de Freud demasiado fuertes, puede haber otras explicaciones del significado de la atribución de estados mentales inconscientes, que involucren supuestos teóricos más débiles: enfoques disposicionales referidos sólo a las semejanzas de conducta [*behavioral*] o circunstanciales, tal vez con la condición adicional de que el estado inconsciente se puede volver consciente bajo ciertas condiciones. Enfoques de este tipo han

sido ofrecidos por filósofos, sin duda. Pero son diferentes al de Freud y, en este caso, sus observaciones sobre lo que él pensaba no pueden ser descartadas tan fácilmente como el *obiter dicta* filosófico de un científico, al hablar de la naturaleza de su actividad profesional inicial.

Vayamos ahora a la segunda de las tres cuestiones planteadas: ya que la concepción de Freud es que tanto los procesos mentales conscientes como los inconscientes son *en sí mismos* físicos —aunque en este momento podemos pensarlos sólo en términos mentales—, puede parecer que él está comprometido con una posición materialista respecto del problema mente-cuerpo. Sin embargo, creo que este no es el caso. Freud aparentemente *aceptaba* una posición materialista, pero esto no se *vincula* a las posiciones que estamos considerando ahora. Todo depende de lo que se diga sobre la conciencia en sí misma. La concepción de Freud sería materialista sólo si la conciencia fuera, también, un fenómeno o un rasgo físico de los procesos cerebrales, que son los procesos mentales conscientes.

Esta es, de hecho, la posición de Freud en el Proyecto, ya que postula una clase especial de neuronas, las neuronas-ω, que se activan en cierto sentido que se identifica con la experiencia consciente. Su afirmación más cuidadosa sobre este punto es la siguiente:

> Deseo decir unas palabras sobre la relación de esta teoría de la conciencia con otras. De acuerdo con una teoría mecanicista avanzada, la conciencia es un mero apéndice de procesos fisiológico-psíquicos y su omisión no produciría ninguna alteración en el pasaje psíquico (de sucesos). De acuerdo con otra teoría, la conciencia es el lado subjetivo de todos los sucesos psíquicos y es por lo tanto inseparable de los procesos fisiológicos mentales. La teoría desarrollada aquí se encuentra entre ambas concepciones. Aquí, la conciencia es el lado subjetivo de una parte de los procesos físicos del sistema nervioso, a saber, de los procesos ω; la omisión de la conciencia no deja a los sucesos psíquicos inalterados, sino que conlleva la omisión de la contribución de ω.[13]

Decir que la conciencia es el *lado subjetivo* de cierto tipo de proceso neurofisiológico no es compatible con el dualismo, aunque tampoco sea correcto llamarlo materialismo. El enfoque parece combinar varios puntos a la vez. (1) Todo proceso mental consciente es[14] un proceso físico, del cual la conciencia es un

aspecto. (2) La conciencia no es un *efecto* del proceso físico. Su existencia no es compatible con la no ocurrencia del proceso físico, ni su ausencia es compatible con la ocurrencia del proceso físico. Esta posición, aunque no ha sido desarrollada, es sutil e interesante. Sus visiones posteriores sobre el tema se encontrarían, probablemente, en su artículo metapsicológico sobre la conciencia, escrito en el mismo año que "El inconsciente".[15] Lamentablemente, este escrito nunca fue publicado y parece haber sido destruido.

Puede ser que Freud haya optado por la teoría del doble-aspecto (si se la puede llamar así); en tal caso, las otras doctrinas que estamos considerando podrían dar una justificación para pensar sobre fenómenos físicos en términos mentales, sin implicar la existencia de procesos no físicos. Sin embargo, sería también posible sostener que la conciencia nos hace advertir procesos psíquicos que son en sí mismos físicos y pueden existir inconscientemente, *sin* dejar que los sucesos de la conciencia sean ellos mismos físicos. En cualquiera de los dos casos, el fundamento para pensar lo inconsciente psíquico en términos de apariencias conscientes es el mismo, y se puede apelar a la analogía con la física.

Pero la posición de que la conciencia es también un proceso físico, o el "lado subjetivo" de ciertos procesos físicos, si bien es oscura, es más interesante desde el punto de vista filosófico. Vale la pena decir algunas palabras sobre cómo se puede construir una concepción de este tipo, ya que atañe a la discusión actual sobre el problema mente-cuerpo. Comúnmente, cuando contrastamos la apariencia fenoménica de algo con su naturaleza objetiva, la explicamos como un *efecto* de la segunda sobre un observador humano. Así, sentimos al hielo frío, en virtud de su efecto sobre nuestro sentido del tacto, y la propiedad física que identificamos con el frío —energía cinética baja— es algo distinto a este efecto sensorial. Si adoptamos la posición correspondiente respecto de la relación entre la conciencia y los procesos cerebrales de los cuales es una apariencia, entonces el estado consciente debería ser descrito como un *efecto* del proceso cerebral sobre la toma de conciencia del sujeto, mientras que el proceso cerebral es algo diferente.[16]

Sin embargo, la concepción de Freud es que el proceso cerebral correspondiente a un estado consciente *no* es algo distinto de la conciencia, que es el darse cuenta de ello. Las cualidades

conscientes no proporcionan una descripción completa del proceso, ya que su naturaleza objetiva es física, y la conciencia es sólo su "lado subjetivo". Pero la conciencia no es un *efecto* del proceso físico, así como la superficie de un objeto no es un efecto del *objeto*. Ni tampoco es una parte que pueda separarse. Los filósofos de la mente no tienen mucho que decir actualmente sobre la hipótesis de que un proceso neural podría aparecerle al sujeto como un proceso consciente, sin producir *efectos* subjetivos, simplemente en virtud de sus propias cualidades subjetivas. Esta me parece una cuestión interesante para indagar.

Vayamos a la última de las tres preguntas planteadas más arriba. ¿Nos da el enfoque de Freud una justificación para *teorizar* sobre el sistema nervioso central en términos mentalistas? Pienso que esta pregunta no ha recibido una respuesta afirmativa en el análisis seguido hasta aquí. Hemos argumentado que la concepción de Freud explica cómo los términos mentales, con o sin la implicación de conciencia pueden, en principio, *referirse* a procesos físicos de los que hoy en día no podemos formarnos una concepción explícita. Esto no significa, sin embargo, que una teoría útil sobre estas cuestiones pueda ser construida utilizando conceptos mentales, como se puede ver si volvemos a la analogía con el uso de imágenes visuales en física.

La visualización es útil al pensar en la estructura molecular o atómica, pero la mayoría de los conceptos cuantitativos importantes de la teoría física o química no se representan visualmente, sino más bien formalmente. Aun nuestra comprensión del mundo visible depende de conceptos tales como peso, energía, impulso [*momentum*], que pueden ser representados visualmente sólo de modo grosero. La teoría física depende del desarrollo de conceptos no fenoménicos.

¿Por qué, entonces, deberíamos esperar que nuestro conocimiento del cerebro progresara al teorizar sobre conceptos fenoménicos de tipo mentalista? Deseos, aversiones, placer y dolor, intenciones, creencias y pensamientos proporcionan, sin lugar a dudas, explicaciones acerca del quehacer de las personas. Pero, ¿hay alguna razón para esperar que el refinamiento y sistematización de estas explicaciones nos dé una teoría sobre cómo opera el sistema nervioso central? Freud argumenta convincentemente que no debemos esperar una teoría completa si nos limitamos a describir las conexiones entre estados que son, de hecho, conscientes. Pero, ¿es suficiente ampliar nuestro cam-

po de investigación para incluir estados psíquicos inconscientes, es decir, análogos en estructura y causas y efectos, a los estados psíquicos conscientes? Esto sería como tratar a la física enteramente en términos de sustancias y fenómenos visibles, más sustancias y fenómenos invisibles, estructuralmente y causalmente análogos. El resultado sería algún tipo de mecanismo. ¿No es el psicologismo una concepción del cerebro análogamente estrecha?

La idea de que podemos esperar descubrir algo sobre el cerebro a través del desarrollo de teorías de la mente en psicología y lingüística fue rescatada recientemente con el mentalismo de Noam Chomsky.[17] No es necesario ofrecer su aporte como una de las justificaciones de la lingüística mentalista, que es hoy en día, después de todo, el único método prometedor disponible para la investigación de cómo funcionan las lenguas naturales. Al menos podemos tratar de descubrir las reglas fonéticas, sintácticas y semánticas que los hablantes parecen seguir. Los lingüistas han logrado un éxito considerable con este modo de describir la gramática. Pero es importante reconocer que si las personas no siguen *conscientemente* determinadas reglas susceptibles de ser enunciadas, en algún área de actividad, no hay ninguna *garantía* de que puedan descubrirse reglas, de las que podamos decir que los hablantes siguen *inconscientemente:* que se comportan *como si* las siguieran, y a las que se ajustan sus juicios respecto del uso correcto e incorrecto.

Esto debería ser evidente para cualquiera que reflexione sobre los fracasos del análisis conceptual en filosofía, ya que el análisis conceptual es un tipo de teoría de la mente que trata de formular reglas para la aplicación de conceptos, cuyos usuarios hablan *como si los siguieran.* Wittgenstein combate enérgicamente en sus *Investigaciones filosóficas* la suposición de que dichas reglas susceptibles de ser enunciadas deban ser descubiertas detrás de nuestras intuiciones de corrección o incorrección en el uso del lenguaje.

Más aún, incluso si una teoría de la mente del tipo *como si* logra dar cuenta de las habilidades humanas o de la competencia en algún dominio, como lo ha hecho la gramática, aparece otro problema: ¿qué significado debemos asignarle a la afirmación de que las personas no hablan meramente *como si* siguieran ciertas reglas, sino que las *siguen* (inconscientemente)? Las razones de esta afirmación no son claras.

Chomsky sugiere, sin tomarlo como una parte central de su concepción que, cuando una teoría de la mente es exitosa en algún área como la gramática, entonces puede llegar a tener significación física. Su comentario sobre esta posibilidad, prudente pero interesante, hacia el final de *Language and Mind*, es el siguiente:

> He estado usando terminología mentalista con toda libertad, pero sin que esto suponga que haya decidido de antemano la cuestión acerca de la posible realización física de los mecanismos abstractos postulados para dar cuenta de los fenómenos del comportamiento o de la adquisición del conocimiento. No estamos obligados, como lo estuvo Descartes, a postular una segunda sustancia cuando discutimos fenómenos que no pueden describirse en términos de lo que él llamaba la materia en movimiento. Tampoco tiene mucho sentido empeñarse en suscitar la cuestión del paralelismo psicofísico, en este contexto. Es una cuestión interesante la de saber si el funcionamiento y la evolución de la mentalidad humana pueden acomodarse dentro del marco de la explicación física, en el sentido actual del término o si, por el contrario, hay principios nuevos, hoy desconocidos, que habrá que invocar, principios que tal vez sólo emerjan a niveles de organización más elevados que los que por el momento permite la investigación física. Podemos, sin embargo, estar bastante seguros de que habrá una explicación física para los fenómenos en cuestión, de darse el caso de que algún día puedan explicarse, por una razón terminológica sin ningún interés, que se reduce al hecho de que sin duda el concepto de "explicación física" será ampliado de modo que en él quepa todo lo que se descubra en ese campo, exactamente lo mismo que ocurrió en el caso de la fuerza gravitatoria y la electromagnética, las partículas sin masa y muchas otras entidades y procesos que habrían representado una ofensa para el sentido común de las generaciones precedentes.[18]

Esto concuerda con la perspectiva que encontramos en Freud, pero da lugar a la pregunta de cómo una teoría de la mente debería desarrollarse antes de que su tema fuera admitido en el mundo físico por derecho propio. Una teoría de la que puede quitarse el carácter mentalista sin pérdida de poder explicativo no es esencialmente mentalista. Se podría construir, por ejemplo, una versión mentalista de la mecánica newtoniana, y describir la mutua atracción de los cuerpos y su obstinación para moverse en línea recta salvo que actúe una fuerza exter-

na (siendo estos estados psíquicos inconscientes, por supuesto). Sin embargo, el contenido explicativo de tal teoría podría ser explicitado en términos más claros, más formales, cuantitativos y no mentalistas.

Por otro lado, si una teoría es esencialmente mentalista dado que su valor explicativo no puede ser recapturado por una versión no antropomórfica, entonces se puede poner en duda que los objetos que describe serán admitidos en el dominio de la física. Esto es porque las descripciones, conexiones y explicaciones mentalistas deben ser entendidas asumiendo, en lo posible, el punto de vista del sujeto de los estados y procesos mentales en cuestión. Aun cuando los estados mentales son inconscientes, la comprensión que nos proporciona dicha teoría requiere que asumamos el punto de vista del sujeto, ya que entendemos la forma de la conexión explicativa entre estados mentales inconscientes y su medio circunstancial y de conducta sólo a través de la imagen de los procesos mentales conscientes, con todas las referencias a significado, intención y percepción de los aspectos involucrados.

Ya que parece ser parte de nuestra idea de un mundo físico que lo que ocurre en él puede ser aprehendido no sólo desde un punto de vista, sino desde indefinidos puntos, ya que su naturaleza objetiva es externa a todo punto de vista, hay una razón para creer que, hasta que no dejemos atrás todos estos rasgos subjetivos, las hipótesis de la psicología mentalista no serán aceptadas como explicaciones físicas. La expectativa de una tal objetivación en la psicología es aún oscura, así como la forma que pueda tomar. Pero este es un tema difícil que no podemos tratar aquí.

Debemos mencionar que algunos psicoanalistas sostuvieron que las teorías de Freud ya estaban muy adelantadas hacia la objetividad: su psicodinámica es impersonal y científica, su terminología antropomórfica es sólo metafórica y no desempeña ningún papel teórico fundamental.[19]

Estos argumentos fueron muy convincentemente discutidos en un artículo de William I. Grossman y Bennett Simon,[20] que es también una excelente guía sobre la bibliografía psicoanalítica acerca de este tema. El pensamiento subjetivo antropomórfico parece indispensable para la comprensión de enunciados como el siguiente:

El médico analista y el yo debilitado del paciente, apoyados en el mundo real exterior, tienen que tomar partido contra los enemigos, es decir, contra las exigencias instintivas del ello y las exigencias de conciencia que sustenta el superyó.[21]

La teoría psicoanalítica tendrá que cambiar mucho antes de poder ser considerada como una parte de la descripción física de la realidad. Y quizás ella y otras teorías mentalistas no alcanzarán nunca el tipo de objetividad necesario para lograr dicho fin. Tal vez, finalmente, las explicaciones físicas de los fenómenos en cuestión no podrán ser alcanzadas con refinamiento y exactitud progresivos en nuestra comprensión mentalista, sino que se relacionarán con las teorías mentalistas de una forma que no podemos percibir.[22] En este momento, así como en 1896, es demasiado temprano para saberlo.

Notas

1. Freud da a menudo la impresión de ser hostil a la filosofía, pero en una carta a Fliess del 1 de enero de 1896, dice: "Veo que usted utiliza el camino sinuoso de la medicina para llegar a su primer ideal, la comprensión fisiológica del hombre, mientras que yo albergo secretamente la esperanza de alcanzar por la misma ruta mi objetivo original, la filosofía. Ya que esta era mi ambición original, antes de saber lo que debería hacer en el mundo" (Carta 39, en *Los orígenes del psicoanálisis* [Basic Books, 1954], p. 141).

2. *Standard Edition of the Complete Psychological Works of Sigmund Freud,* comp. James Strachey (Hogarth Press, 1961 [en lo que sigue, *E.S.*], vol. 5, esp. 612-13, 615-16.

3. *E.S.*, vol. 14, esp. 166-71.

4. *E.S.*, vol. 23, Cap. 4 a 8, esp. 157-60 y 196-97. [*Esquema del psicoanálisis*. Barcelona, Paidós, 1982.]

5. *E.S.*, vol. 23, esp. 282-83 y 285-86.

6. *E.S.*, vol. 23, 196-97. [Paidós, pp. 88-90.]

7. Esta es la posición expresada por Freud en su monografía *Sobre la afasia*: "Es probable que la cadena de sucesos fisiológicos en el sistema nervioso no esté en conexión causal con los sucesos psíquicos. Los sucesos fisiológicos no cesan en cuanto los sucesos psíquicos comienzan; por el contrario, la cadena fisiológica continúa. Lo que ocurre es simplemente que, después de cierto tiempo, cada una (o alguna) de sus conexiones tiene un fenómeno psíquico que le corresponde. De acuerdo con ello, lo psíquico es un proceso paralelo al fisiológico: 'un concomitante dependiente' " (*E.S.*, vol. 14, 207).

8. *E. S.*, vol. 14, 171.

9. De modo similar, es posible hablar de sonidos tan débiles o tan agudos que no pueden ser oídos por ningún organismo porque tenemos una teoría física del sonido.

10. En ningún caso cree que estos fundamentos operen como premisas de una *inferencia* consciente, sin embargo: "(Sin duda sería psicológicamente más correcto ponerlo de este modo: que sin ninguna reflexión especial les atribuimos a los otros nuestra propia constitución y por lo tanto nuestra conciencia, y que esta identificación es una condición *sine qua non* de nuestra comprensión)" (*E.S.*, vol. 14, 169).

11. Puede ser que Freud se apresure al hacer esta suposición. Su motivo principal para rechazar la extensión de la analogía con otras mentes a la atribución de conciencia al inconsciente es que "la conciencia, sobre la que su poseedor no sabe nada, es algo muy diferente a la conciencia que le pertenece a otra persona, y es cuestionable que tal conciencia, que carece de su característica más importante, merezca discusión" (*E.S.*, vol. 14, 170). Pero, naturalmente, si el inconsciente fuera conciencia de *sí mismo*, entonces tendría un "poseedor" distinto del sujeto de la conciencia habitual en la misma persona, y sería sólo este último inconsciente de estos estados conscientes del inconsciente. Sin embargo, Freud ofrece otras razones en contra de su carácter consciente, es decir, problemas de incompatibilidad, peculiaridad, incoherencia e indeterminación del número de sujetos requerido para acomodar todos los otros estados conscientemente en una forma más o menos unificada.

12. Véase, sin embargo, *Más allá del principio de placer*: "No tenemos que sentirnos muy afectados al juzgar nuestra especulación sobre los instintos de vida y muerte por el hecho de que muchos procesos oscuros y confusos ocurran en ella, tales como que un instinto es acarreado por otro, o que un instinto se aleja del ego hacia un objeto, etc. Esto se debe a que tenemos que operar con términos científicos, es decir, con el lenguaje figurativo, peculiar de la psicología (o, más precisamente, de la psicología profunda). Si no, no podríamos describir los procesos en cuestión, y seguramente no nos habríamos podido percatar de ellos. Probablemente, las deficiencias de nuestra descripción desaparecerían si estuviéramos ya en condiciones de reemplazar los términos psicológicos con términos fisiológicos o químicos. Es verdad que estos son también sólo parte de un lenguaje figurativo, pero es un lenguaje con el que hace mucho que estamos familiarizados y que es, quizás, más simple" (*E.S.*, vol. 18, 60).

13. *E.S.*, vol. 1, 311.

14. Esta identificación se encuentra explícitamente formulada en el siguiente pasaje: "Por lo tanto, reunimos coraje para suponer que hay un tercer sistema de neuronas (podríamos llamarlo ω, quizás) que se excita conjuntamente con la percepción, mas no con la reproducción, y cuyos estados de excitación dan lugar a las diferentes [cualidades]; son, se diría, *sensaciones conscientes* (*E.S.*, vol. I, 309).

15. Véase la introducción del compilador a los artículos sobre metapsicología, *E.S.*, vol. 14, 105-7.

16. Véase Saul A. Kripke, *Naming and Necessity* (Harvard University Press, 1980), esp. pp. 144-55.

17. Véase Noam Chomsky, *Aspects of the Theory of Syntax* (M.I.T. Press, 1965), p. 193. [*Aspectos de la teoría de la sintaxis*. Barcelona, Gedisa, 1999.]

18. Noam Chomsky, *Language and Mind* (Harcourt, 1968). [*El lenguaje y el entendimiento*, Planeta, 1986, p. 162.]

19. Por ejemplo, H.Hartmann, E. Kris y R.M. Loewenstein, "Comments on the Formation of Psychic Structure", en *The Psychoanalytic Study of the Child* 2 (1946), pp. 11-38. También en *Psychological Issues* 14 (International Universities Press, 1964), pp. 27-55. A veces se argumenta como una crítica que las teorías de Freud son fundamentalmente mecanicistas. Véase, por ejemplo, A.C. MacIntyre, *The Unconscious* (Routledge, 1958), p.22: "Aunque Freud abandonó finalmente y terminantemente el intento de lograr una explicación neurofisiológica... mi crítica —la más importante de esta parte del argumento— es que Freud conservó esta concepción de la mente como una pieza de un mecanismo, y que simplemente describió en términos psicológicos lo que originariamente pretendía ser una teoría neurológica".

20. "Antropomorfismo: Motivo, significado y causalidad en la teoría psicoanalítica", en *The Psychoanalytic Study of the Child* 24 (1969), pp.78-111.

21. *E.S.*, vol. 23, 173 [Paidós, p. 52]

22. Se encuentra una opinión similar en Donald Davidson, "Mental Events" (1970), reimpreso en Davidson, *Essays on Actions and Events* (Oxford University Press, 1980).

2

La revolución permanente de Freud*

Este artículo fue elaborado a partir de una reseña de The Mind and Its Depths, *de Richard Wollheim (Harvard University Press, 1993) y de* Freud and His Critics, *de Paul Robinson (University of California Press, 1993), publicada originalmente en el volumen del 12 de mayo de 1994, de* The New York Review of Books. *El addendum es una respuesta a los comentarios de la reseña de Adolf Grünbaum, publicados en el volumen del 11 de agosto de 1994.*

Los grandes revolucionarios intelectuales cambian el modo como pensamos. Plantean nuevas preguntas y proponen nuevos métodos para contestarlas. Y no podemos desaprender tales formas de pensamiento al descubrir errores de razonamiento por parte de sus creadores, salvo que estemos convencidos de que esos pensamientos son idénticos a los errores. Hay algo extraño en los debates recientes sobre la evidencia en la cual Freud basó sus teorías. Su influencia no es como la de un físico que afirma haber descubierto una partícula hasta entonces no observada con un experimento que otros consideran defectuoso. Cualquiera fuera el futuro del psicoanálisis como una forma especial de terapia, la influencia de Freud no me parece más apta para ser erradicada de la conciencia moderna que la de, por ejemplo, Hobbes o Descartes. Estos pensadores tienen un efecto demasiado profundo como para ser captado por una

* Reimpreso con autorización de *The New York Review of Books.* Copyright 1994 por Nyrev, Inc.

serie de hipótesis determinadas, un efecto que ni siquiera desaparecería si, por estar en una ola de "Eurofobia", dejáramos de leer sus escritos.

I

La interpretación correcta de la influencia de Freud, y el modo como debemos evaluarla, es un tema común a los dos libros de esta reseña; *The Mind and Its Depths* nos proporciona, además, un ejemplo de dicha influencia. Se trata de una colección de ensayos sobre el arte, la moral y la mente que escribió Wollheim durante el período en que también publicó *The Thread of Life*[1] y *Painting as an Art,*[2] los cuales abarcan contenidos comunes con estos ensayos. En *The Mind and Its Depths* nos encontramos con uno de los pensadores contemporáneos más psicoanalíticos. Wollheim tiene una gran percepción de la realidad e influencia del inconsciente, y del impacto de la sexualidad infantil sobre toda la vida mental posterior.

Wollheim sostiene que lo que Freud alcanzó fue una gran ampliación del *insight* psicológico arraigado en la psicología de sentido común, utilizando algunos de sus conceptos, pero yendo mucho más lejos. Los *insights* psicológicos no son poco frecuentes, ya que nos pasamos la vida tratando de entendernos a nosotros mismos, pero el objetivo y el carácter imaginativo de los métodos de Freud para entender crean un problema especial de interpretación y evaluación.

El problema es el siguiente: como Wollheim observa, las explicaciones de sentido común son una forma de comprender "desde adentro"; aun cuando nos permiten penetrar en la mente de otros, dependen, en parte, de la comprensión de uno mismo, ya que interpretan al otro como un otro yo [*self*]. Para entender los pensamientos, sentimientos o comportamiento del otro, debemos darle un sentido —aunque sea un sentido irracional— de su punto de vista, utilizando nuestro propio punto de vista como un recurso imaginativo. La imaginación nos permite darles un sentido interno a las creencias, emociones y objetivos que no compartimos: ver cómo se relacionan entre sí, a fin de hacer inteligible la conducta del otro. Pero la extensión de Freud de este tipo de *insight* a los pensamientos, motivos y fantasías, y a la mente del niño, amenaza con quitarles a los

conceptos psicológicos ordinarios —como creencia, deseo y anhelo— el soporte familiar que se funda en la experiencia y comprensión de la vida cotidiana.

Algunos, como Sartre, encontraban contradictoria la idea de un pensamiento del cual no tenemos conciencia, pero Freud podía responder a esta objeción. El era un materialista sofisticado en metafísica; creía que aun los procesos mentales conscientes eran también eventos físicos en el cerebro, aunque no supiéramos nada acerca de su carácter físico. Si esto es así, entonces tiene sentido preguntarnos si acaso no podría haber otros procesos cerebrales, análogos a los conscientes en la estructura física, y con causas y efectos visiblemente psicológicos, pero no conscientes. Su realidad sería física, aun cuando podríamos conocerlos sólo a través de sus manifestaciones psicológicas, así como, hoy en día, podemos referirnos (salvo pocas excepciones) a los procesos cerebrales conscientes sólo en términos psicológicos.[3] De modo que el problema principal del inconsciente no es metafísico. El problema es, más bien, si la evidencia apoya una extensión tan amplia de conceptos mentales al inconsciente por medio de la analogía, trayendo consigo el crecimiento concomitante de la explicación psicológica.

Con la psicología de sentido común podemos identificar las experiencias o deliberaciones que dieron origen a una creencia, o las emociones expresadas por una reacción determinada, o los propósitos o valores que subyacen a un curso de conducta. Los juicios causales de esta clase son mayormente automáticos: llenan nuestra vida y nuestras relaciones con los demás, y los defendemos por su utilidad, aunque también nos puedan conducir al error. Cuando interpretamos a los demás de este modo, dándole sentido a su punto de vista, no estamos meramente imaginando cosas, como cuando vemos animales en las nubes o le atribuimos malicia a una tostadora defectuosa. Más bien estamos tratando de entender, dentro de los límites de la psicología no científica, lo que realmente hace que la gente actúe, y a menudo esperamos que el propio autoconocimiento de la persona nos ratifique.

Freud extendió el alcance de este tipo de explicación a niveles sin precedentes, abarcando no sólo olvidos y deslices, sino también bromas, sueños, síntomas neuróticos, y la subestructura de la vida erótica y de los lazos familiares, incursionando

en la moral, la política, el arte y la religión. Para algunos de estos fenómenos no había ninguna comprensión psicológica adecuada; para otros, Freud propuso agregarles un nivel de comprensión más profundo que el que la psicología consciente podía proporcionar. En muchos casos, lo hizo a través de *insights* que los demás podían evaluar fácilmente desde su propia experiencia y observación, pero que apelaban a la evidencia alcanzada por medio del muy poco usual tratamiento psicoanalítico, evidencia que sólo podían comprender aquellos que estaban familiarizados con el proceso.

Sin embargo, todo el sistema siguió siendo psicológico, en el sentido que Wollheim especifica: trataba de lograr una comprensión de los seres humanos "desde adentro", de modo que podamos ponernos en su lugar y darles sentido a sus síntomas y respuestas, atribuyéndoles creencias, deseos, sentimientos y percepciones, con la diferencia 'e que estos eran aspectos de puntos de vista de los que ellos no se daban cuenta conscientemente. ¿Qué razón hay para pensar que una extensión tan amplia de la interpretación psicológica no es meramente una fantasía, tal como ver animales en las nubes, y no conocimiento genuino?

Es un hecho conocido que la gente puede no darse cuenta de sus motivos verdaderos, y que a menudo entendemos a los demás mejor de lo que ellos se comprenden a sí mismos. (Esto lo sabe cualquiera que haya escuchado con desconcierto a una persona citar nombres de famosos para hacerse notar, en lo que creía ser una conversación amable.) Pero Freud llevó esta idea tan lejos que no podía defenderla apelando sencillamente al sentido común. El insistía en el carácter científico de sus descubrimientos y el apoyo que les daba la observación clínica, que significaba no sólo experimentos controlados, sino los datos que aparecen en el proceso de análisis. La interacción permanente y única entre el analista y el paciente neurótico le proporcionaba una evidencia mucho más amplia y sistemática para la interpretación de lo que es accesible a alguien que meramente observa los síntomas del paciente, incluso alguien que conoce al paciente muy bien.

Creo que podemos entender mejor la confianza de Freud como la creencia en que la exposición, en gran cantidad de casos, a las minuciosas descripciones de la experiencia que surgen en el análisis, permiten ver una coherencia psicológica más pro-

funda en fenómenos que, tomados aisladamente, parecen carentes de significado o inexplicables. No hay duda de que tal coherencia puede conducir a errores. Nos podemos imaginar, por ejemplo, que un trastorno mental causado por una droga puede producir complicados patrones de pensamiento y sentimiento que parecen apuntar a una causa psicológica, y no química. En general, es importante tener en cuenta la posibilidad real de que un síndrome que tiene "sentido" psicológico pueda tener una explicación puramente fisiológica. Pero esto sólo quiere decir que la evidencia psicoanalítica, como la mayoría de las evidencias, no es definitiva.

Los comentarios más directos de Richard Wollheim sobre este tema aparecen en un ensayo titulado "Desire, Belief, and Professor Grünbaum's Freud". Se trata de una respuesta al libro *The Foundations of Psychoanalysis* de Adolf Grünbaum,[4] un libro que toma en serio las pretensiones científicas del psicoanálisis, pero interpretándolas de un modo curiosamente superficial, dejando de lado el carácter peculiar interno del *insight* psicológico. Hacia el final del ensayo, Wollheim observa en forma expresiva: "Si... la teoría psicoanalítica es una extensión de la psicología del sentido común, tal vez deberíamos empezar por preguntarnos: ¿Cómo se pone a prueba la psicología de sentido común?" Considero que la respuesta es que la evidencia de la psicología de sentido común es compleja y dispersa, en lugar de ser el resultado de experimentos controlados. Wollheim le achaca a Grünbaum tener una concepción empobrecida de cómo trabaja la explicación psicológica, y dejar de lado el papel esencial de la estructura psicológica —es decir, la necesidad de usar categorías psicológicas de *cierto* tipo— tanto en la comprensión psicoanalítica como en la de sentido común.

> Si debemos apoyar lo que el paciente dice o hace en la hipótesis que estamos considerando, de modo que podamos decir que la hipótesis fue comprobada en el diván, entonces el material del paciente deberá ser subsumido, en la mayoría de las circunstancias, bajo categorías que derivan del psicoanálisis... Al decir o hacer lo que el paciente hace, él tiene que ser identificado como si presentara *material anal* en grado masivo; recurriendo a *fantasías de omnipotencia*; *agrediendo* o *fragmentando* o *idealizando* la interpretación del analista;

actuando; etc. En otras palabras, el material del paciente debe ser subsumido bajo categorías de transferencia: esto es, categorías que captan lo que la persona está haciendo *vis-à-vis* la situación analítica, tal como él la fantasea. (p. 108)

Cada una de estas categorías o hipótesis que explican la conducta del paciente tiene que encontrar su apoyo empírico en innumerables aplicaciones a otros pacientes en otros contextos. En otras palabras, el psicoanálisis hace uso de una compleja red de interpretaciones, tal como hace la psicología de sentido común cuando nos permite entender las reacciones de alguien, relacionándolas con una red de deseos, creencias, emociones, memorias, obsesiones, inhibiciones, valores e identificaciones. Y Wollheim observa que aun donde Grünbaum propone una alternativa más o menos intuitiva [de sentido común] a la explicación psicoanalítica, a saber, que la evidencia clínica aparente a favor del psicoanálisis es el resultado de la sugestión por parte del analista, no ve la necesidad de hacer comprensible psicológicamente la dinámica de dicha sugestión, o de explicar los procesos mentales a través de los que opera. La hipótesis de "sugestión", después de todo, es una explicación *psicológica* alternativa, y debe ser evaluada con los mismos estándares que se aplican a la explicación que está tratando de refutar.

¿Cómo sabemos si una explicación psicológica es correcta? Aunque no necesitamos un análisis estadístico para probar que alguien se puso un saco porque tenía frío, podemos imaginarnos fácilmente un nimio experimento psicológico, que confirma estadísticamente la probabilidad de un nexo causal en tales casos. Pero cuanto más interesante es el caso, más difícil es reproducirlo. Por ejemplo, ¿con qué evidencia experimental podríamos responder la pregunta de por qué Mijaíl Gorbachov comenzó a desmantelar el imperio soviético? Todo lo que podamos decir sobre ello dependerá de la aplicación de métodos generales para la interpretación motivacional de la conducta intencional en las circunstancias específicas del caso.

Gran parte de la vida mental humana consiste en sucesos complejos con múltiples causas y condiciones de fondo, que no se repetirán nunca. Si tratamos de entender la vida real, es inútil requerir experimentos repetibles y con controles estrictos. (El mismo problema surge con respecto a la explicación

histórica, ya que los acontecimientos históricos son aun menos repetibles.) Esto no quiere decir que la explicación sea imposible, sólo que no puede ser encontrada con los métodos propios de la física de partículas, de investigación del cáncer, o el estudio de los reflejos. Podemos no ser capaces de hacer experimentos controlados, pero podemos todavía tratar de darle un sentido interno a lo que la gente hace, a la luz de sus circunstancias, tratando de ver cómo podría parecer justificado desde su punto de vista, una técnica que ha sido apoyada por su utilidad en innumerables otros casos, ninguno de ellos idéntico al otro.

Las explicaciones que se refieren a los procesos mentales inconscientes deberían ser evaluadas con el mismo estándar. Puede haber algunas explicaciones psicoanalíticas tan simples pasibles de ser comprobadas por experimentos o análisis estadísticos, pero la mayoría no son de este tipo: son, más bien, aplicaciones del *insight* psicoanalítico en circunstancias sumamente específicas, que van más allá de los límites de la conciencia. Cuando nos encontramos con una conjetura tan brillante y rica como la que Freud atribuye al olvido de una palabra en una cita en latín, a la ansiedad del sujeto de que su amante pueda estar embarazada,[5] es completamente imposible darle una confirmación estadística y, simplemente, debemos decidir si esto es una extensión intuitivamente verosímil de una estructura general de explicación que está bien fundamentada en alguna otra parte, o si es más admisible que las otras alternativas, incluyendo aquella de que no hay ninguna explicación psicológica.

El mismo problema surge para propósitos más generales, como la hipótesis de Freud de que en casos de melancolía, la pérdida de un objeto (tal como la partida de un amante) se transforma en una pérdida del yo (el sentir que uno no vale nada, o es desdeñable), a través de la identificación del yo con el objeto del amor, por el cual fue abandonado; el odio hacia uno mismo que resulta puede ser entendido como un abuso del objeto internalizado, por parte del yo. No parece haber modo de evaluar esta propuesta en forma experimental; sin embargo es una hipótesis empírica sobre un proceso psíquico que, ciertamente, parece dar cuenta de lo que ocurre en algunos casos de depresión aguda.[6]

Para la mayoría de los que creen en la realidad de la depresión y del inconsciente, hayan sido psicoanalizados o no, la

creencia se basa no en la confianza ciega en la autoridad de los psicoanalistas y sus observaciones clínicas, sino en la utilidad evidente de la visión freudiana para entenderse a uno mismo y a los demás, sobre todo, la vida erótica, los dramas familiares y lo que Freud llamaba la "psicopatología de la vida cotidiana". Cosas que, de otro modo, nos sorprenderían, no nos sorprenden; la conducta o los sentimientos que, de otro modo, parecerían irracionales, se vuelven comprensibles. Un día nos sentimos desgraciados todo el tiempo, y luego descubrimos que era el aniversario de la muerte de alguien importante para nosotros, nos enojamos reiteradamente con ciertas mujeres de nuestro ámbito profesional, y luego nos damos cuenta de que nuestro enojo remite a una pelea que tuvimos en la infancia con nuestra madre. Al final, si debemos creer que Freud se estaba acercando a la verdad, deberíamos poder hacer nosotros cierto uso de su teoría. Ya que los experimentos controlados y repetibles no son posibles aquí, la clase de comprensión interna, característica del psicoanálisis, debe descansar en la confirmación dispersa, pero acumulativa en la vida, que apoya nuestros juicios psicológicos familiares.

La cuestión no es si Freud estaba exactamente en lo cierto, o si no se pueden hacer críticas a algunas de sus historias de casos, o si los tipos de explicación que él introdujo amplían sustancialmente la comprensión de nosotros mismos y de los demás que nos proporciona la psicología de sentido común. Creo que la profunda transformación freudiana de nuestra concepción moderna del yo evidencia la validez de su intento de extender lo psicológico más allá de su base consciente. Esto, a su vez, hace creíble que los analistas que exploren mucho más profundamente y descubran mucho más material para la interpretación puedan desarrollar *insights* más extensos y sistemáticos del mismo tipo. Muchos de nosotros *sentimos* como si durante gran parte del tiempo, la conciencia nos revelara sólo la superficie de nuestra mente y, para muchos, esta sensación se ve confirmada en sus sueños.

II

La opinión de Grünbaum está bastante distante de la perspectiva básica freudiana, que es un rasgo tan familiar de la

cultura moderna. Wollheim, en cambio, se encuentra cómodo con un rico trasfondo de categorías psicoanalíticas, algunas desarrolladas por Melanie Klein. El pone un énfasis especial e interesante en la fenomenología del inconsciente, es decir, los sentimientos subjetivos y fantasías del inconsciente que, sostiene, son esenciales para su poder explicativo. Por ejemplo, Wollheim propone la tesis del "yo corpóreo", según el cual concebimos, cuando somos muy pequeños,

> estados mentales con el modelo de entidades corpóreas... un pensamiento como un trozo de comida en la boca, o como heces, y concebimos aceptar el pensamiento como tragar el trozo de comida, o rechazar el pensamiento, como excretar las heces. (p. 87)

Ya que los mecanismos de defensa de Freud incorporan estas fantasías, no son fuerzas abstractas, sino que tienen un carácter subjetivo muy específico (aunque inconsciente).

Es difícil para un amateur evaluar tales afirmaciones; aun cuando el tema en más familiar, la dificultad permanece. La comprensión psicoanalítica de la moral es un ejemplo interesante. La historia de la formación del superyó, "la versión interna del padre en el drama edípico", ha pasado a la psicología popular.

Pero Wollheim se enfrenta a esta concepción de la moral, y sostiene con Melanie Klein "que los mandatos/entredichos o detonantes de figuras internas que yacen en el núcleo del yo desempeñan, en el mejor de los casos, un papel poco fiable y en el peor de los casos, en la vida moral". En otras palabras, concibe al superyó como un objeto incorporado, amenazador, castigador y ajeno, y no como un yo mejor con el cual la persona llega a identificarse. En lugar de fundamentar el desarrollo moral en la internalización del padre reprobador, Wollheim, siguiendo a Klein, lo basa en la reconciliación de sentimientos muy anteriores. Luego del primer año de vida, el niño descubre que el objeto que él ama y el objeto que él odia son, ambos, su madre; el niño trata de reparar, preservar o resucitar el objeto amado, que él ha dañado con su fantasía omnipotente. Con el tiempo, el niño llegará a integrar al yo bueno y al malo, al yo que ama y al que odia.

Wollheim describe este proceso, que es considerablemente más complejo que este breve bosquejo, en un ensayo llamado

"The Good Self and the Bad Self", donde compara a Klein con el filósofo F.H. Bradley,[7] quien también sostenía que la moral debe ofrecer una forma de autorrealización armónica. Wollheim es un naturalista en lo moral, en el sentido de que cree que nuestras afirmaciones sobre nosotros mismos deben estar enraizadas en nuestros sentimientos más tempranos y más básicos. Nos urge a evitar "la fantasía de que la moral marca el lugar donde los seres humanos descartan la naturaleza humana". Esto parece querer decir que no debemos tratar de entender la moral como una trascendencia radical de las necesidades y deseos infantiles y de las estructuras mentales establecidas antes de los dos años de edad. Salvo que la moral esté enraizada en dichas necesidades y deseos, será superficial.

No puedo evaluar la teoría kleiniana del desarrollo infantil, ni siquiera aunque yo sea una instancia de ella, pero aparte de eso, no veo cómo una teoría de este tipo podría explicar más que los comienzos del complejo sistema de restricciones sobre la agresión y autointerés, el reconocimiento de obligaciones formales y de los derechos y reclamos de los demás, que dan lugar a una moral completamente desarrollada. Lo mismo puede decirse de la más familiar teoría freudiana del superyó. Aun si comienza con un drama familiar, la moral en sentido estricto necesita formas de pensamiento que son mucho más impersonales que el miedo, el amor, o la identificación con determinados "objetos" internos o externos, sean el padre o la madre. La moral trata de proveernos con estándares objetivos en el ámbito de la conducta, que nos permitan justificarnos mutuamente y estar de acuerdo sobre qué es lo que se debe hacer.[8]

La actitud de Wollheim sobre la moral es mucho más radical que la de Freud, y, en espíritu, más cercana a la de Nietzsche. En un ensayo llamado "Crime, Punishment, and 'Pale Criminality' ", aplica su naturalismo de una manera sorprendente, sugiriendo que los seres humanos tienen una disposición "ligada a lo que es lo más profundo en nosotros, a hacer lo que está prohibido, y a hacerlo por esa razón"; y que si esto es así, entonces el sistema de justicia penal se basa en una comprensión de la naturaleza humana equivocada: lo que está prohibido es más atractivo. Muchas instituciones sociales, tales como el derecho penal, "implican una psicología determinada", dice Wollheim, y si la psicología está equivocada respecto de los seres huma-

nos, "entonces la institución está equivocada, ya que, en lugar de facilitar el autoconocimiento, lo impide".

Esto, dice, origina una pregunta más:

> ¿Cuánta seguridad podemos esperar de una conducta criminal? ¿Cuánta protección podemos reclamar con justicia de aquellos con quienes compartimos nuestra psicología? No la criminalidad, sino las semillas de la criminalidad están presentes en todos nosotros de un modo identificable que la ciencia de la mente pueda esclarecer, ¿hasta dónde tenemos el derecho de alejarla de nosotros?... ¿Hasta qué punto deberíamos estar preparados para sacrificar la seguridad en pro del autoconocimiento, en caso de tener necesidad de ello? (p. 130)

Pero, ¿cómo surge esa necesidad? Tal vez la disposición humana para hacer lo que está prohibido cause que algunas prohibiciones aumenten el atractivo del crimen, pero esto sólo quiere decir que si queremos seguridad, debemos tener esto en cuenta al diseñar el derecho penal. ¿Dónde está el sacrificio de autoconocimiento aquí? Sospecho que Wollheim está hablando de algo más profundo que el autoconocimiento. El reclama una suerte de autoafirmación que es incompatible con el rechazo de cualquier cosa que esté en el núcleo del yo; esto explica su descrédito de la moral convencional.

El arte es diferente. Wollheim escribe:

> Me parece natural pensar que el arte está arraigado más profundamente en la naturaleza humana que la moral, y me sorprende que los filósofos no le den importancia al hecho de que, aunque el buen arte es más agradable que el malo, la gente virtuosa no goza de esta ventaja respecto de aquellos que son valorados primariamente por su encanto, o su alegría, o su dulzura, o su violencia. (p. x)

En varios ensayos, Wollheim analiza la expresión, percepción y estilo artístico y, aquí también, se compromete con el naturalismo psicológico: "las amplias características del arte, incluyendo la expresividad, originan arte exterior [*outside art*]". Wollheim enfatiza la importancia psicológica de la "proyección", la instilación de nuestras percepciones del mundo exterior con los estados mentales internos, que él cree que se encuentra en el corazón de la respuesta estética. El arte le da sentido a nues-

tra vida con su expresividad, que crea una concordancia entre nuestros sentimientos profundos e interiores y nuestras percepciones externas, y nos hace sentirnos a gusto en el mundo. Nuevamente, aquí, así como en el análisis de la moral, surge la pregunta de si una teoría tan personal no hace a la respuesta estética demasiado subjetiva e idiosincrásica; pero esto dependerá de cuán objetivas consideramos a estas respuestas y cuán universales son las estructuras mentales que responden, según su teoría, a las proyecciones originadas por el arte. Wollheim mismo piensa que hay una naturaleza humana universal, que presupone todo el arte "o, por lo menos, todo el gran arte".

Tanto el arte como la naturaleza pueden ser objetivos de sentimientos proyectados: tanto un paisaje real como uno pintado pueden verse como melancolía, por ejemplo. Pero, con una obra de arte, Wollheim cree que la verdadera respuesta está determinada por la intención del artista: la obra fue creada para evocar ciertas proyecciones y, si esta intención es satisfecha, el espectador la verá en ella. Wollheim rechaza la opinión de que "la crítica tiene la libertad de proyectar en una obra de arte lo que desee". Sostiene que la propia respuesta proyectiva del artista a la obra que está creando —en su doble rol de creador y espectador— desempeña un papel esencial en el proceso creativo: "el hecho central del arte [es] que se trata de una manifestación intencional de la mente".

Painting as an Art contiene muchas aplicaciones de esta idea. Por ejemplo, Wollheim le atribuye el efecto sobrenatural del gran paisaje *The Large Enclosure near Dresden*, de Caspar David Friedrich, a la presencia de un "espectador invisible en la pintura", cuyo punto de vista no tiene una relación natural con el paisaje. Cuando miramos el cuadro nos vamos identificando con este espectador incorpóreo dentro de él; nos encontramos tanto ante el cuadro como flotando sobre el paisaje, que se extiende debajo del espectador invisible, y nos vemos arrastrados a la actitud de la contemplación indiferente a la vez que absorta de la naturaleza, que expresa el pietismo decimonónico temprano de Friedrich. (La única vez que vi este cuadro, me produjo una extraña desorientación espiritual, como si hubiera perdido mi yo y estuviera viendo el mundo *sub specie aeternitatis*. Por ello, la explicación de Wollheim me pareció convincente.)

Una cuestión diferente es cuánto conocimiento previo necesitamos para ver la obra tal como pretende el artista que la veamos. Percibir el significado de una obra de arte generalmente requiere más conocimiento que el que necesitamos para encontrar hermoso un paisaje natural: aun percibir las propiedades formales y no representativas requieren que la obra "se perciba como parte de una tradición estética". Pero si la obra tiene éxito, entenderla no consistirá en un acto deliberado de inferencia e interpretación, sino simplemente en percibirla como se pretende que uno la perciba. La esencia de la comprensión estética se encuentra en la experiencia, no en la teoría. No se encuentra ni en la explicación social o económica, ni en la decodificación simbólica.

El atractivo de la posición de Wollheim se encuentra en su insistencia en que lo que es importante en una obra de arte es directamente perceptible en ella. Más difícil de aceptar es su afirmación de que el fin de la percepción estética es experimentar la proyección del artista. Esto implica que una obra de arte no puede significar más que lo que el artista quiso que significara, o sabía que significaba, y que la interpretación histórica y simbólica y formal no puede producir una comprensión perceptiva de la obra que la interpreta como algo más que un producto de los propósitos del artista. Una posición contraria a la de Wollheim podría incluir la posibilidad de que el significado de una obra pueda desarrollarse en el tiempo, a medida que el trasfondo de la interpretación cambia. A mí me atrae el énfasis de Wollheim en la experiencia estética. Sin embargo, es posible que su teoría vaya demasiado lejos al reducir la estética a la psicología. El papel dominante que le da tanto a la intención como a la proyección significa que la obra de arte no es una creación autónoma, cuyas características estéticas trascienden ambos hechos psicológicos. Dudo que ni siquiera los grandes artistas sepan lo que están haciendo, en todos los aspectos, aun cuando no sea accidental que sus obras causen los efectos que de hecho producen.

LECTURA CULTURA SIN EGOISMOS
BIBLIOTECA
www.lecturasinegoismo.com

III

Los escritos de Wollheim, diversos y originales sobre muchos temas diferentes, ejemplifican la influencia de los modos

de pensamiento psicoanalíticos más allá de la terapia. Paul Robinson enfatiza tal influencia en su enérgica respuesta a tres comentaristas recientes de Freud: Frank Sulloway, Jeffrey Masson y Adolf Grünbaum.[9] *Freud and His Critics* contiene reflexiones interesantes y generales sobre la significación de reacciones recientes a Freud, así como una disección de estos tres escritores. Robinson argumenta convincentemente que ninguno de los tres logra dar cuenta de sus más pretenciosas afirmaciones sobre Freud y, además, que pierden de vista el significado real de la contribución intelectual del creador del psicoanálisis.

Sulloway es un historiador que se presenta como un nuevo exégeta de Freud, no como uno de sus críticos. Pero Robinson cree que él disminuye el interés y la originalidad de las ideas de Freud, al exagerar su contenido biológico a expensas del psicológico. Debería ser evidente a cualquier lector de Freud que su convicción freudiana de que la mente, que es una función del cerebro, es un producto de la evolución biológica, y que su estructura está sujeta a influencias evolutivas, y Sulloway presenta una detallada explicación del trasfondo biológico y neurológico del desarrollo intelectual de la teoría freudiana. Pero tratar esto como la esencia de una comprensión freudiana de la mente es leer a Freud en forma demasiado reduccionista. Puede ser que la mente sea un producto biológico, pero los conceptos biológicos nos proporcionan sólo una comprensión superficial de su contenido y forma de trabajar. Sulloway magnifica la influencia que ejercieron sobre Freud las excéntricas teorías de su confidente Wilhelm Fliess, aunque hayan dejado sus huellas. Es verdad que Freud mantenía correspondencia con Fliess sobre los efectos de los calendarios biológicos o del sentido del olfato o la evolución de la postura erecta sobre la mente. Pero no es posible, según Robinson, reemplazar al *insight* psicológico —del tipo "desde adentro" que ocupó realmente al talento de Freud— con hipótesis de este tipo.

Más adelante, Robinson se dirige al insigne Masson, que acusa a Freud de ser un mentiroso y un cobarde, porque abandona su afirmación de que sus primeros pacientes habían sido víctimas de abuso sexual real en la infancia, explicando sus síntomas con la teoría de la sexualidad infantil y de la fantasía. Masson cree que Freud perdió de este modo la oportunidad de ser un gran luchador en contra del abuso infantil, y que

su teoría se convirtió en un medio de acusar a víctimas inocentes. Robinson demuestra que las afirmaciones de Masson a favor de que los pacientes habían sido, de hecho, víctimas de abuso sexual, son meras aseveraciones no fundamentadas. Y agrega: "La más poderosa objeción a la tesis de Masson de la cobardía moral, es que Freud abandonó la teoría de la seducción sólo para adoptar una idea que ofendía aun más los prejuicios de su cultura, a saber, la teoría de la sexualidad infantil". Freud siempre reconoció la existencia del abuso en niños: en todo caso, dudaba de su extensión. Pero la afirmación de que algunas acusaciones de abuso de la seducción infantil son producto de la fantasía provoca gran resistencia, no sólo por parte de Masson. Creo que esta insistencia en la inocencia del niño y en la conducta perversa de los adultos revela sentimientos más profundos, que emergen en una mezcla emocionalmente rica de sensualidad y violencia. La actual obsesión popular sobre el abuso sexual infantil se debe, en gran parte, a esos sentimientos.[10] Robinson identifica a Masson de forma convincente con el nuevo puritanismo que emergió a comienzos de la década de 1980 como una reacción a la década de expansión sexual de 1960. La visión de Masson del sexo como carente de alegría y cargado de agresión pertenece a una tendencia antiliberal que ha ido creciendo recientemente. Calza bien en la perspectiva de los que ven a la pornografía sólo como un instrumento para la dominación de la mujer. Freud es un blanco perfecto para este tipo de enemigos del autoconocimiento.

La crítica que hace Robinson a Grünbaum se dirige a la importancia que él le asigna al éxito terapéutico como la base empírica sobre la que se apoyan las teorías freudianas. Freud niega en varias partes que la terapia eficaz sea una prueba última de sus teorías.[11] La teoría de la represión tiene poder explicativo y tiene apoyo en una variedad de fuentes. Estoy de acuerdo con Robinson en que, con imaginación y agudeza, es posible, e incluso a veces fácil, extrapolar a partir de lo consciente y familiar, y descubrir explicaciones psicológicas inconscientes en casos individuales complejos en los que la verificación estadística es imposible.

La evidencia a favor de esta afirmación sólo se obtiene con ejemplos particulares, que no nos dejan ninguna alternativa creíble. Aunque haya quienes puedan considerar convincentes las famosas historias de casos de Freud, creo que son demasia-

do complejas y ambiguas para servir a este propósito. No es sorprendente que hayan generado tantas controversias históricas e interpretativas. La mejor evidencia para los escépticos es de menor escala. En una carta a Freud, Einstein escribió una vez:

> Hasta hace poco tiempo podía comprender el poder especulativo de su pensamiento y su enorme influencia sobre la *Weltanschauung* de la era actual, sin estar en posición de formarme una opinión definitiva sobre el grado de verdad que tiene. Sin embargo, hace poco tuve la oportunidad de escuchar algunas instancias que, en sí mismas, no son tan importantes, pero que, a mi juicio, excluyen cualquier otra interpretación que la que otorga la teoría de la represión.[12]

Pero, ¿cómo podía saberlo Einstein? ¿Qué es esto de una concatenación de circunstancias que "excluyan cualquier otra interpretación"?

Sin saber a qué instancias se refería Einstein, puedo citar un episodio que yo mismo presencié. En una fiesta, un hombre mayor, de buena posición social, que había pasado su vida como investigador privado, sin cargos académicos, desafió a un psiquiatra que se encontraba presente a explicar por qué, cada vez que escuchaba las noticias en la radio, se dormía en el momento en que pasaban el informe de la bolsa. El psiquiatra le contestó que, tal vez, este hecho reflejaba sentimientos complicados respecto de su padre. "¡Mi *padre*!" dijo el hombre, incrédulo, "¡mi padre se murió hace cincuenta años!" Luego se habló de otra cosa pero, al día siguiente, el hombre llamó al psiquiatra para contarle que, más tarde, en ese mismo día, le había venido a la memoria el momento en que, en su juventud, él se había negado a participar en la empresa familiar, y su padre le había obligado a prometerle que al menos iba a escuchar todos los días el informe de la bolsa por radio.

Mucha gente ha presenciado ejemplos tan obvios como este, aunque la mayoría de ellos no son tan esquemáticos, y el material que surge en el psicoanálisis es mucho más complejo y extraño. Grünbaum es reacio al uso del *insight* psicológico para extender formas de explicación familiares y básicas a situaciones radicalmente no familiares. Pero esto puede deberse en parte a que él mismo tiene una imaginación psicológica más

bien rígida. Como señala Robinson, cuando Grünbaum propone una interpretación diferente de la de Freud, para los mismos datos, su interpretación sale mal parada. (Lo mismo que su idea de que un desliz lingüístico que puede deberse a un pensamiento encubierto pero consciente es "el hombre que aleja su mirada del pecho descubierto de una madre dando de mamar, 'discúlpeme, tengo que tomar un pecho de aire fresco'[get a *breast* of *flesh* air]".)

Robinson concluye con una excelente afirmación del legado intelectual de Freud, y que estos críticos no han podido reconocer:

> Es la principal fuente de nuestra inclinación moderna buscar significados debajo de la superficie de la conducta: estar siempre alertas buscando el significado "real" (y presumiblemente oculto) de nuestras acciones. También nos inspira la creencia de que los misterios del presente se volverán más transparentes si podemos rastrearlos hasta sus orígenes en el pasado, incluso en el pasado más remoto... Y, finalmente, nos despertó una mayor sensibilidad para lo erótico, sobre todo, su presencia en escena... donde las generaciones pasadas dejaron de buscarlo. (pp. 270-71)

Sin embargo, el libro tiene el defecto de una gran confusión, que aflora ocasionalmente como, por ejemplo, en el siguiente pasaje: "No se puede saber nunca, a partir de la lectura de Sulloway, Masson y Grünbaum, que muchos de sus contemporáneos tenían profundas dudas sobre la ciencia, la objetividad, la verdad y la posibilidad de adquirir un conocimiento estable e incontrovertible del yo y de la sociedad". Robinson describe a los tres escritores como "positivistas", por su apego inocente a ideas pasadas de moda de verdad y objetividad, y considera esta misma actitud como un rechazo de la concepción de Freud:

> El modernismo implicaba una pérdida de confianza en la estabilidad y transparencia del yo. También implicaba el reconocimiento de que todo el conocimiento humano es subjetivo e indeterminado. La teoría del inconsciente de Freud, que rechaza que el yo sea consciente incluso de sus propias ideas, era la articulación más poderosa de esta sensibilidad modernista. (p. 16)

Robinson se refiere al subjetivismo fácil que hoy en día arruina muchas de las humanidades y ciencias sociales. De acuerdo

con esta posición, cualquiera que piense que algunas preguntas tienen respuestas verdaderas y falsas, que pueden ser refutadas por evidencia y argumentos, son verdaderos hombres de las cavernas.

Robinson le atribuye equivocadamente esta posición a Freud. Hay una gran diferencia entre sostener que no somos transparentes para nosotros mismos y que debemos descubrir nuestra verdadera naturaleza mental con métodos de investigación difíciles e indirectos, y sostener que no hay tal cosa como la verdad o la objetividad. El inconsciente no elimina la objetividad, aunque la haga más difícil de alcanzar. En el sentido de Robinson, Freud era un "positivista" y un convencido de la búsqueda de la verdad por medio de la evaluación adecuada de evidencia.

Esto no implica que lo que creemos verdadero sea inmune a la revisión a la luz de evidencia o argumentación posterior, ni implica que todo puede ser conocido. Pero sí implica que, aunque, al igual que cualquier ciencia, la psicología se apoya en la imaginación para enmarcar sus hipótesis, su objetivo es descubrir verdades sobre la mente humana. También implica que, si todo lo que Freud logró fue desarrollar una nueva manera de hablar o ver las cosas, fracasó. Esto es lo que él quería decir en su reiterada insistencia de que lo que él hacía era ciencia. Freud habría estado encantado de debatir con Grünbaum, y no habría tolerado los ataques sobre la objetividad, que Robinson describe en forma deprimente como "el intelectual más visible de la actualidad". No ayudamos a Freud si lo defendemos apelando a esta perspectiva, y menos aún, atribuyéndosela.

Addendum

Un problema fundamental para que progrese esta disputa es que no hay acuerdo sobre lo que debe ser considerado la contribución distintiva de Freud, específicamente, para el propósito de determinar su validez, valor o grado de apoyo empírico. Grünbaum lo identifica con una serie de principios psicológicos. Yo lo identifico con una determinada manera de entender, que se manifiesta en innumerables interpretaciones y explicaciones individuales. Ambos estamos de acuerdo en que las hipótesis

psicoanalíticas son causales y requieren confirmación empírica, pero no estamos de acuerto en cuanto al tipo de evidencia que es más importante. No niego la importancia de posibles futuros tests epidemiológicos y experimentales del tipo que él considera crucial. Mi principal desacuerdo con Grünbaum es sobre si hay ahora, que no existen tales experimentos, razones sustanciales para creer en el inconsciente y en las explicaciones psicoanalíticas que se refieren a él, de la clase de razones que eran también accesibles a Freud.

No creo que la cuestión crucial para evaluar la herencia de Freud es si él tenía razón (para tomar el ejemplo de Grünbaum) al proponer que la homosexualidad reprimida es una condición necesaria causal de delirios paranoicos. Incluso aparte de la probabilidad de que los delirios paranoicos a menudo tienen causas no psicológicas, esta generalización universal parece de lo más inaceptable, y no hay duda de que los escritos de Freud están llenos de hipótesis generales igualmente falsas. Pero el núcleo de su contribución se encuentra en otra parte, en la forma de la explicación que depende no de la aplicación de leyes específicamente psicoanalíticas, sino de la extensión de las formas familiares de explicación más allá de su dominio racional tradicional.

Pongo el acento en la continuidad entre las explicaciones psicoanalíticas y las de sentido común, porque creo que el psicoanálisis puede tomar prestada la evidencia empírica para sus fundamentaciones generales más importantes desde la confirmación del sistema de la explicación psicológica ordinaria en la vida cotidiana. (Aquí sigo no sólo a Richard Wollheim, sino a otros dos filósofos, Donald Davidson y James Hopkins, quienes desarrollan este punto de vista mucho más profundamente de lo que yo puedo hacerlo aquí.[13]) La psicología de sentido común no depende sólo de generalizaciones causales del tipo citado por Grünbaum (que la gente se enoja con insultos, que buenas noticias producen alegría, etc.); una lista de leyes de este tipo no llegarían demasiado lejos en nuestro conocimiento de las personas. Mucho más importante es el esquema general con el que tratamos de entender a los demás como seres más o menos racionales —cada uno con un sistema complejo de creencias, supuestos, preferencias, deseos, valores, objetivos y disposiciones para hacer inferencias— e interpretamos su conducta como dotada de propósitos e intenciones a la luz de dichas condiciones.

El principio causal fundamental de la psicología de sentido común es que, en la mayoría de los casos, podemos descubrir condiciones causalmente relevantes (que hacen una *diferencia* justamente en el sentido de Grünbaum) para una acción o pensamiento o emoción humana, al ubicarla en una interpretación racionalmente coherente de la persona en tanto sujeto intencional, al ver cómo está *justificada* de algún modo, desde el punto de vista de la persona. La interpretación revela causación, porque esa es la clase de sistema que es el ser humano. Y este principio está tan bien apoyado en innumerables casos simples, en los cuales puede ser confirmado por posibilidades de predicción y control, que estamos totalmente justificados al aplicar el mismo principio para identificar causas psicológicas como casos únicos e irrepetibles, tratando de darles un sentido intencional y dotado de propósitos.

Esto es lo que quiero decir con credibilidad intuitiva, y se aplica necesariamente en primera instancia para explicaciones específicas, más bien que a principios generales. Creo que la esencia del método freudiano fue extender el alcance de este sistema explicativo a áreas de la conducta y del sentimiento humano, donde antes parecía no haber ningún sentido. De este modo, aumentó nuestro conocimiento de la influencia de la mente, pero la confirmación va de lo particular a lo general: la teoría general de la represión y desarrollo psicosexual tiene que apoyarse en instancias individuales, más bien que a la inversa.

No poseo la clase de experiencia que me permitiría formarme un juicio sobre los conflictos entre diferentes teorías posfreudianas en el nivel de los principios generales. Pero el método general freudiano de extender el esquema familiar interpretativo de explicación al inconsciente en los casos particulares, método del cual depende la evidencia de dichas teorías, es algo que todos deberíamos poder confirmar con nuestra propia experiencia: se trata simplemente de darle sentido a la conducta irracional, no intencional o involuntaria, cuando cabe en el mismo tipo de patrón tan familiar de la psicología ordinaria, completando algunos de los baches con pensamientos o deseos que el sujeto no conoce conscientemente.

El caso del que se duerme durante el informe de la bolsa, que ejemplifica innumerables casos, debería ilustrar este punto. La reacción de Grünbaum a este caso muestra cuán lejos

estamos uno del otro. Primero dice que Freud se equivocó al pensar que hay una tendencia a olvidar experiencias dolorosas. Estoy totalmente de acuerdo en que esto no es una tendencia universal. En efecto, creo lo contrario. Pero el punto es irrelevante, porque ninguna generalización de esta clase cabe como base para la explicación, que depende solamente de los particulares de cada caso. En segundo lugar, Grünbaum dice que yo no doy lugar a "la rápida creación, por parte del psiquiatra, de una clase-mental [*mind-set*] en el individuo, cuando le dice, sin *ninguna* información adicional, que el que se queda dormido durante el informe de la bolsa probablemente exprese sentimientos difíciles respecto de su padre". Grünbaum parece no haber captado la significación de la evidencia crucial —que el psiquiatra conoce— de la ambivalencia experimentada por el individuo respecto de la riqueza que había heredado, al dedicarse a la extraña vida de investigador privado, fuera de los círculos académicos. Esto es lo que originó la interpretación, ya que, a la luz de este trasfondo, los síntomas que eran inexplicables *adquieren un sentido expresivo.* La memoria recuperada, evocada vívidamente por la sugerencia de interpretación, cabe en el lugar como otra pieza del mismo rompecabezas, y refuerza fuertemente el "sentido" del síntoma, aunque esto es obviamente sólo el comienzo de la historia: no conocemos qué mezcla de desafío, temor, amor y culpa caracteriza los sentimientos del individuo hacia su padre.

Pero Grünbaum no toma nada de esto. Su referencia a una "creación de una clase mental" implica, creo, que probablemente el recuerdo era falso e inculcado por la sugestión del psiquiatra. Pero preferir esta alternativa a otra que le da sentido al síntoma revela, creo, tanto la profunda alergia de Grünbaum a admitir motivos inconscientes, como su celeridad para apelar a la sugestión como una explicación alternativa para todo propósito ante la menor excusa. No sé si el síntoma cesó luego de este episodio, pero no dudo de que si hubiera ocurrido así, Grünbaum lo habría atribuido también a la sugestión.

Al apelar a la profunda influencia de la ideas freudianas sobre la autoconciencia moderna, quise decir que todos nosotros utilizamos constantemente estas formas de comprensión, que la experiencia nos presenta continuamente en las circunstancias adecuadas. Evidentemente, esto no es verdad en el caso de Grünbaum; tal vez él lo considere como una gran vulgariza-

ción. Sin embargo, trataré de contestar esta última pregunta.[14] Las supersticiones religiosas, las patrañas étnicas, la brujería y la esclavitud son ejemplos muy diferentes de errores arraigados, pero quiero centrarme en la brujería, ya que se la cita a menudo en estos debates. Ninguno de nosotros cree en la brujería, pero la pregunta interesante es por qué no. ¿Podría Grünbaum apelar a experimentos controlados que establecieran que maleficios hechos por brujas diplomadas no tienen ningún efecto estadístico sobre el porcentaje de muerte de sus víctimas? Yo no estaría interesado en estos datos, aun si existieran. Mi motivo para rechazar la brujería, aun cuando innumerables personas creen en ella, es que sabemos por razones mucho más generales que el mundo no funciona de ese modo. No hay fuerzas sobrenaturales que puedan invocarse al clavar agujas en muñecos. Esto deriva simplemente de nuestro conocimiento general de física y biología.

Pero el psicoanálisis no es una teoría de fuerzas sobrenaturales. Es una ampliación de la explicación psicológica a otros fenómenos *dentro* del dominio de su aplicación original, es decir, las vidas de los seres humanos dotados de mente. Esto no es de ningún modo incompatible con el resto de nuestro conocimiento científico de cómo funcionan las cosas. Es parte de nuestra idea del orden natural de que la conducta de la gente está influida por su condición mental: que la influencia debe ser mayor y más variada de lo que pensábamos originalmente no debería sorprendernos.[16]

Un comentario final sobre la eficacia terapéutica, que es indirectamente relevante para el tema de la confirmación. No he tenido experiencia de primera mano de psicoanálisis o psicoterapia, pero conozco mucha gente que la ha tenido. Aunque no sé si el psicoanálisis es más o menos efectivo para eliminar síntomas no deseados que una medicación o terapia conductista, por ejemplo, estoy seguro de que tiene una *clase* de efecto sobre los pacientes diferente de otras formas más "externas" de tratamiento. Mi observación es que el psicoanálisis puede otorgar una valiosa herramienta de autoconocimiento, que es profunda aunque esencialmente perceptiva y no teórica, y que este autoconocimiento, sea que cure las neurosis o no, puede ser *utilizado* para anticipar, identificar y manejar formas de irracionalidad que, de otro modo, actuarían de manera perjudicial o inhabilitadora. También permite una respuesta más sutil a

la irracionalidad neurótica de los demás, gracias a la ampliación de la imaginación psicológica. Por esta razón, creo que va a sobrevivir al desarrollo de los tratamientos sintomáticos más simples, aunque, debido a su costo en tiempo y dinero, su forma tradicional será siempre una opción para una pequeña minoría.

Notas

1. Harvard University Press, 1984.
2. Princeton University Press, 1987.
3. Este tema aparece en varios puntos en los escritos de Freud, incluyendo el "Proyecto para una psicología científica" (1895), "El inconsciente" (1915) y *Esquema del psicoanálisis* (1938).
4. University of California Press, 1984. Grünbaum sostiene que Freud apoya su [*case for*] teoría de la represión en la mayor efectividad terapéutica del psicoanálisis para el tratamiento de neurosis, y en que tal evidencia no es accesible. Su interpretación de Freud, y de la evidencia, clínica y extraclínica, ha sido extensamente criticada, sobre todo por David Sachs, en "In Fairness to Freud", *Philosophical Review* 98, nº 3 (Julio de 1989), y por varios comentaristas en *Behavioral and Brain Sciences* 9 (Junio de 1986). Más recientemente, publicó *Validation in the Clinical theory of Psychoanalysis* (International Universities Press, 1993), una discusión sobre estos temas, que incluye material nuevo así como versiones de ensayos publicados anteriormente, algunos anteriores a *The Foundations of Psychoanalysis.*
5. *The Psychopathology of Everyday Life* (1901), capítulo 2.
6. Véase "Mourning and Melancholia" (1917). Freud enfatiza que su propuesta se aplica sólo a algunos casos de melancolía, en tanto que otros parecen ser de origen somático.
7. Bradley es el tema del primer libro de Wollheim, *F.H. Bradley* (Penguin, 1959).
8. Una compilación reciente sobre la obra de Wollheim contiene dos análisis valiosos sobre este tema, por Marcia Cavell y Samuel Scheffler. Ambos arguyen o que una concepción más objetiva y menos "auto" centrada de la moral puede ser compatible con una teoría psicoanalítica del desarrollo moral. Véase Jim Hopkins y Anthony Savile (comps.), *Psychoanalysis, Mind and Art: Perspectives on Richard Wollheim* (Blackwell, 1992). Scheffler desarrolla sus ideas sobre el tema en *Human Morality* (Oxford University Press, 1992), y Cavell expone su posición en *The Psychoanalytic Mind: From Freud to Philosophy* (Harvard University Press, 1993), que incluye una crítica aceptable a la teoría freudiana de la moral.
9. Además de analizar el libro de Grünbaum de 1984, analiza *Freud, Biologist of the Mind*, de Sulloway (Basic Books, 1979) y *The Assault on Truth: Freud's Suppression of the Seduction Theory*, de Masson (Farrar, Straus y Giroux, 1984), tratando también algunas de las respuestas que recibieron dichas obras.

10. Y ha causado algunas terribles persecuciones a inocentes. Por ejemplo, en el *New York Times* del 21 de noviembre de 1993 (p. 29), se informa sobre la expulsión de un maestro de escuela dominical, acusado de violaciones y torturas espeluznantes, sobre la base de testimonios obtenidos de sus alumnos, que en ese momento eran niños de tres o cuatro años de edad.

11. Por ejemplo, en sus *Introductory Lectures* de 1917 (*E.S.*, vol. 16, 255).

12. Citado en Ernest Jones, *The Life and Work of Sigmund Freud* (Basic Books, 1957), vol. 3, p. 203.

13. Véanse Hopkins, "Introduction: Philosophy and Psychoanalysis", en Hopkins y Wollheim (comps.), *Philosophical Essays on Freud* (Cambridge University Press, 1982); "Paradoxes of Irrationality", de Davidson, en el mismo volumen; y "Epistemology and Depth Psychology: Critical Notes on *The Foundations of Psychoanalysis* en S. Clark y C. Wright (comps.), *Psychoanalysis, Mind and Science* (Blackwell, 1988).

14. Grünbaum se preguntó: "Si la influencia cultural penetrante fuera evidencia de validez, entonces las supersticiones religiosas y los estereotipos o patrañas étnicos, que son mucho más dominantes que las ideas freudianas, así como fueron antes la brujería y la esclavitud, deberían poseer un alto grado de validez. ¿Les aplica Nagel este criterio cultural también? Si no lo hace, ¿por qué?

15. Incidentalmente, el mayor alcance de la mente también se revela de una forma totalmente diferente: con el descubrimiento reciente de que procesos fisiológicos, por lo común completamente involuntarios, como los latidos del corazón o la presión alta, pueden ser controlados, con el entrenamiento adecuado, de forma consciente y voluntaria.

3

Wittgenstein: El conflicto egocéntrico*

Este ensayo fue originalmente una reseña de The False Prison: A Study of the Development of Wittgenstein's Philosophy, *vol. 2 (Oxford University Press, 1988), de David Pears. La principal pregunta de los intérpretes de Wittgenstein es cómo entender el papel que él les asignó a las condiciones empíricas del significado, condiciones de acuerdo natural y modo de vida común, dentro del cual deben hacerse todas las justificaciones, correcciones y críticas. En sus escritos se sugiere una interpretación "idealista": estos modos de vida son* lo dado, *sobre los que se asienta todo el resto. Pero esto da una concepción de la verdad tan obviamente falsa en diversos ámbitos que creo que no puede ser una interpretación correcta, aunque no parece que haya otra alternativa. Es también una teoría filosófica sobre la naturaleza de la realidad demasiado general como para atribuírsela al segundo Wittgenstein. Cualquiera haya sido su respuesta, la lectura de Wittgenstein de la crítica kantiana del pensamiento filosófico, que trata de trascender las condiciones de su propia significación, sigue siendo la concepción filosófica más profunda del siglo* XX.

En la década de 1950, cuando yo era estudiante en Cornell, esta era la única universidad norteamericana donde se estudiaba intensamente la última obra de Wittgenstein. El murió en 1951, y sus *Investigaciones filosóficas* fueron publicadas en 1953. Recuerdo aquellos días previos a la fotocopia, en que

* Reimpreso con autorización de *The London Review of Books*, 18 de mayo de 1989.

compartí con algunos compañeros el trabajo de copiar a máquina *El libro azul* y *El libro marrón* con múltiples papeles carbónicos, a partir de una serie de textos pre-*Investigaciones*, de mediados de la década de 1930, que había en Ithaca y que circulaban en samizdat hasta que se publicaron finalmente en 1960 como una parte de flujo creciente de los volúmenes del *Nachlass*.

El ambiente que rodeaba el estudio de Wittgenstein era emocionante y, al mismo tiempo, sofocante: la filosofía era agonía, y era necesario sumergirse tan profundamente en los problemas que uno apenas podía respirar. La superficialidad representaba el mayor peligro; no se podía alcanzar nada sin luchar, ni acercándose a los problemas ni entendiendo lo que Wittgenstein decía. Sobre todo, había una sensación de que era extremadamente difícil —o imposible— expresar la verdad, dado el fracaso del propio Wittgenstein en publicar el enorme material que escribió luego de volver a la filosofía a la edad de cuarenta, de modo que, salvo en las *Investigaciones*, él se encuentra en una relación especial de responsabilidad menor frente a ella.

Todavía pienso que esta actitud es básicamente correcta. No hay ningún modo de acercarse a Wittgenstein que no sea metiéndose hasta la coronilla en problemas aparentemente insolubles, para luego ver si los lugares que él sugiere para ubicarnos nos permiten avanzar. Hoy en día el nombre de Wittgenstein se usa siempre como el símbolo de un relativismo conceptual despreocupado, y las referencias al Argumento del lenguaje privado y a las Formas de vida son referencias casi tan comunes como las del Principio de indeterminación de Heisenberg. Pero se trata de una curiosidad cultural: la obra de Wittgenstein es apenas más accesible hoy en día de lo que era hace treinta y cinco años. Es demasiado fácil adoptar las soluciones sin entender los problemas. Como observa David Pears, "cuando leemos alguno de los análisis de Wittgenstein sobre ilusiones filosóficas, hay dos cosas que no podemos tener en nuestra mente simultáneamente: su éxito en disiparlas, y la profundidad y dificultad de los problemas que las originaron". En el segundo volumen de su excelente estudio en dos volúmenes, Pears trata de hacer la obra posterior de Wittgenstein accesible a través de los problemas profundos tradicionales a los que él respondía: su respuesta era tratar de encontrar una salida a ellos. Esto significa que el libro es difícil. Pears no es ni en tempera-

mento ni en estilo un sufriente, pero es fiel a su tema, y, aunque haya cosas con las que no estemos de acuerdo, le da al lector un sentido de la profundidad y el radicalismo de Wittgenstein.

De acuerdo con el título del libro, Pears traza el desarrollo del pensamiento de Wittgenstein mayormente con relación a un tema que ha sido central en la filosofía occidental desde Descartes, y que Wittgenstein encontró en Schopenhauer: la sospecha de que nos encontramos atrapados en nuestras mentes, y que nada de lo que podamos hacer con el lenguaje, pensamiento, imaginación o percepción, nos permitirá salir de ellas. Descartes pensó que este problema se refería a qué podemos conocer. Suponía que podríamos, por lo menos, concebir un mundo externo: el problema era si podíamos saber algo de él o no. Pero, en el desarrollo de la filosofía moderna a partir de Kant, este problema evolucionó como una duda más radical sobre qué podemos pensar. Aun si hay un mundo más allá de nuestra mente, no parece haber ninguna manera de que nuestra experiencia haga contacto con él, o que pueda representarlo tal como es en sí mismo, de modo que el alcance de nuestro pensamiento se limita a nuestras experiencias actuales o posibles, incluyendo experiencias de percepción "externas".

Sin embargo, esta posición no es estable. Si ni siquiera nuestros pensamientos pueden ir más allá de nuestra mente, la idea de un mundo inalcanzable más allá de ella es una ilusión conceptual: no es posible tal pensamiento; cualquier cosa que consideremos como el pensamiento de lo que es impensable por nosotros debe ser otra cosa, o es una farsa. No podemos decir que pensamos el mundo *sólo* como se nos aparece, porque la contrastación correspondiente no tendría sentido.

En el primer libro de Wittgenstein, el *Tractatus Logico-Philosophicus* (1921), este resultado se encuentra en la posición de que el solipsismo coincide con el realismo puro. Todo en el mundo es igualmente real —desde mis impresiones sensoriales hasta las estrellas—, pero el mundo sigue siendo *mi* mundo. Esto se muestra en el hecho de que, independientemente de cuán objetivamente yo pueda describir el mundo (incluyendo los vacíos que debo llenar con cosas que no sé, y aun si esa descripción objetiva no le proporciona nada a la persona particular que soy yo), puedo agregar en forma redundante: "Y soy yo quien dice y piensa esto". "Todo en el mundo" es una

expresión de mi lenguaje. Pero no puedo pensar que estoy atrapado, ya que esto requeriría algo imposible: formar el pensamiento de lo que me es imposible pensar.

El *Tractatus* ofrecía una teoría general del lenguaje y del pensamiento, y de su relación con el resto de la realidad. Los escritos posteriores de Wittgenstein rechazan la posibilidad de alguna teoría de este tipo, a favor de una descripción fragmentaria de ejemplos a partir de la gran variedad de actividades lingüísticas y mentales que los humanos pueden tener.[1] Pero el conflicto egocéntrico sigue siendo una preocupación central. Wittgenstein saca a luz y trata siempre de disolver la duda sobre la adecuación de nuestro lenguaje y pensamiento para alcanzar el mundo que trata de comprender. Los detalles de su respuesta dependen de una clase particular de pensamiento en cuestión, sea el mundo físico, otras mentes, matemáticas, acción, sensación, o el significado. Pero la estrategia general es clarificar una ininteligibilidad en esas dudas que es bastante simple: si son expresables, deben ser falsas. Si yo digo: "¿Cómo es posible que si pronuncio meramente la palabra 'Aristóteles' pueda referirme a *Aristóteles*?" parece que me olvidé del comienzo de la oración cuando llego al final, ya que la segunda ocurrencia del nombre tiene que referirse a Aristóteles para que la pregunta tenga sentido. No puedo aprehenderlo a Aristóteles mismo, por medio de alguna referencia supralingüística, contra la que pueden criticarse las condiciones naturales ordinarias de aplicación del nombre como inadecuado para su propósito.

El mismo problema de la inteligibilidad aparece si pregunto cómo, al usar la palabra "más", puedo aprehender la operación de la *suma*, que es una función definida sobre todos los infinitos pares posibles de números, de los que sólo puedo obtener una pequeña muestra, o la pregunta de si mi atribución de dolor a otra persona realmente se refiere al *dolor*, en tanto opuesto a alguna otra sensación, o a nada en absoluto. ¿Cómo debemos entender la segunda parte de estas preguntas si las preguntas son serias?

Según Wittgenstein, el problema filosófico surge cada vez que arrancamos el lenguaje, las imágenes mentales u otros vehículos de pensamiento de las condiciones que les dan significación, como si realmente pudiéramos captar la realidad sólo trascendiendo esos límites. Entonces usamos esos "pensamien-

tos" trascendentes para cuestionar la adecuación de los pensamientos mundanos a sus condiciones mundanas para expresar lo que pretenden expresar. Nuestro pensamiento y lenguaje actuales parecen deficientes con referencia a un ideal inalcanzable. Pears cita un pasaje de "Notes for Lectures on Private Experience and Sense Data",[2] que concluye con: "No es lo que me reprochan, como si dijeran: '¡en su lenguaje usted está solamente *hablando*!' " Pears agrega: "El lenguaje no puede absorber bien su tema", y "si pensamos en él como algo sublime, hablado desde ningún lugar, las exigencias que ejercerá sobre el mundo serán demasiado exorbitantes como para poder ser satisfechas y el resultado será el escepticismo".

Gran parte del libro de Pears está dedicado al desarrollo a partir de 1929 de la crítica de Wittgenstein al escepticismo sobre la posibilidad de concebir por parte de otra mente y no la propia (una clase de solipsismo lógico, en tanto opuesto al epistemológico) y, menos radicalmente, sobre la posibilidad de atribuir sensaciones como las propias a otros sujetos de experiencia a los cuales uno está limitado (la privacidad, aparentemente esencial, del lenguaje de sensaciones).

Para entender a Wittgenstein es necesario poder tomar estos problemas en serio. Tratamos de entender nuestro propio sistema de lenguaje y pensamiento desde adentro, y nos vemos forzados hacia la conclusión de que en ese sistema otras "personas" son meros objetos de nuestra percepción externa y que las experiencias internas, en el sentido en el que nosotros las tenemos, no pueden ser atribuidas inteligiblemente a esos otros. No puedo imaginarme que *siento* el dolor de otro, ya que, si pudiera, no sería el dolor de él sino el mío. Así, aplicada a los otros, la idea del dolor no es realmente la de un sentimiento sino de un síndrome externamente observable, compuesto por circunstancias y conducta. Mi idea de sentimientos reales, y de un yo real, no se puede aplicar inteligiblemente más que a mi propio caso. El uso de las mismas palabras para referirse a los "otros" es una clase de juego de palabras. Este es el concepto de mente que Wittgenstein combatía en la década de 1930.

Pears muestra cómo el ataque al solipsismo prefigura el ataque posterior al lenguaje privado. En ambos casos, Wittgenstein da vuelta la tortilla sobre el escéptico, argumentando que los pensamientos que tenemos cada uno de nosotros sobre nuestro propio caso —"Yo" o "el modo como veo el rojo"— que, en el

pensamiento escéptico se suponía que daban un punto de partida seguro para tratar de generalizar y para fracasar en generalizar, se vuelven ininteligibles si tratamos de entenderlos aisladamente de sus análogos de segunda y tercera persona: la identificación que hace otra gente de una persona y del carácter de sus experiencias.

¿Qué sería más perfectamente seguro y transparente, se podría pensar, que mi idea de mí mismo —de mi yo— y mi uso de la palabra "yo" para designarla? Si siento dolor en mi mano, sé de inmediato y sin necesidad de observaciones de ningún tipo que soy yo el que siente dolor: de modo que mi identificación subjetiva de mí mismo está completamente autocontenida y es independiente de la identificación de mi cuerpo o de otra cosa en el mundo. Pero Wittgenstein argumenta que esto es una ilusión, producida al abstraer el lenguaje de las conexiones entre la experiencia y sus manifestaciones externas que nos permiten atribuirles experiencias a otros y, por lo tanto, nos permiten aprender a hacerlo con nosotros mismos.

La posibilidad del pensamiento de mí mismo depende de hechos familiares y naturales. Como dice Pears, "es un hecho familiar que, cuando a una persona se le lastima la mano derecha, hay una línea que corre hacia la base [*seat*] de su conciencia, y sale hacia su boca. Y dice, 'yo tengo dolor' y, aunque la palabra 'yo', tal como él la usa, no significa 'este cuerpo', sí presupone la integridad de esta línea personal. Tiene que haber también una conexión que vuelva de la boca hacia la base de la conciencia de dolor a la parte lastimada del cuerpo". Y mi infalibilidad para identificarme como el sujeto de mi dolor —que parece mostrar un yo puro, independiente del cuerpo— puede de hecho explicarse con las palabras de Wittgenstein, de este modo: "El hombre que grita de dolor, o dice que tiene dolor, *no elige la boca que lo dice*". La identificación infalible de mí mismo como el sujeto de una experiencia es meramente una extensión lingüística de ello, más bien que la identificación con un "yo". En otras palabras, no es la aplicación de un concepto puro de primera persona.

Wittgenstein no era conductista, tal como aclara Pears. Pero pensaba que lo que Pears llama "línea personal" desempeña un papel indispensable en tanto condición de posibilidad de los conceptos mentales y del lenguaje mental, y que no podemos

usar ese lenguaje para describir circunstancias alejadas demasiado radicalmente de dichas condiciones. Esto es verdad no sólo para la idea de cómo veo el rojo, o cómo siento el dolor. Se oponía a lo que Pears llama el "Dogma del sentido de largo alcance", que nos inclina a pensar que las palabras llevan su significado a todas partes.

Pasando del yo al contenido de experiencia, veamos el problema filosófico irresistible del espectro invertido: aun si el solipsismo es falso, ¿qué evidencia tengo de que los colores se les aparecen a los demás tal como se me aparecen a mí? Ya que no puedo mirar en su mente, nunca tendré una razón para creer que otros no puedan ver la sangre tal como yo veo el pasto, y viceversa, incluso si no hubiera *ninguna* diferencia anatómica o conductual entre nosotros. Pero la inteligibilidad de esta hipótesis parece depender de alguna base alternativa de significación para la atribución de cualidades sensoriales a los otros, que sobrevive intacta aun si se supone que las conexiones habituales con la anatomía, circunstancias y capacidad discriminatoria y lingüística son evidentemente inviables. ¿Y qué podría ser?

Es claro que se supone que hay una proyección directa en la mente del otro de la conexión inmediata entre la palabra "rojo" y una cualidad sensorial particular que experimento en mi propio caso, sin tener que apoyarme en ninguna de estas conexiones externas. Miro el frasco de ketchup y me pregunto: "¿Lo que este frasco contiene es lo mismo que *esto*?"

Deberíamos poder prever la respuesta de Wittgenstein. "Lo mismo que *esto*" es un intento de estirar un segmento del lenguaje más allá de sus condiciones usuales de aplicación: para expresar una sobrecomparación con referencia a la cual los criterios habituales de comparación pueden considerarse inadecuados. Pero, por más urgente que parezca, no se le ha dado una base alternativa de significación. Como dice Wittgenstein, "cuando se dice: 'o bien él tiene esta experiencia, o no' lo que nos ocurre en primer lugar es una imagen que, por sí misma, parece convertir el sentido de las expresiones en *imposible de ser confundido*. 'Ahora usted sabe lo que está en cuestión', nos gustaría decir. Y esto es, precisamente, lo que la imagen no dice".

No sólo eso, sino también para la atribución de cualidades sensoriales a nuestras *propias* experiencias, que parecen ser,

desde adentro, totalmente independientes de las conexiones hacia afuera, con las circunstancias, conducta y objetos estándar de la percepción cuya significación depende, según Wittgenstein, de la existencia de dichas conexiones en general. Si digo "Por 'rojo' quiero decir lo mismo que *esto*", estoy suponiendo que mi idea subjetiva de un tipo de sensación está bien definida independientemente de su conexión con otra cosa. Pero, de hecho, mi identificación de color de primera persona en tanto lo mismo o lo diferente fue aprendida junto a la capacidad de discriminación que aparece primero en relación con objetos estándar externos y que puede ser identificada por otros hablantes que me enseñan el lenguaje.

Podemos imaginarnos que esto es sólo un artificio que establece una conexión entre las palabras y los tipos de sensación en la mente, una conexión en la cual esas relaciones externas no tienen parte. La respuesta a esta propuesta nos lleva al argumento del lenguaje privado, y al muy debatido tema de las reglas.

Para que una palabra sea más que un ruido o un garabato o un parpadeo en la mente —para que signifique algo—, debe haber una distinción entre su aplicación correcta e incorrecta. ¿Qué determina esta distinción? No puede ser lo que el hablante dice, ya que entonces no estaría *diciendo* nada. La palabra debe llevar algo más. ¿Qué es, y cómo el hablante hace una conexión con ello cuando usa la palabra con ese significado?

Esta es una pregunta enormemente difícil. Wittgenstein observó que las condiciones de fondo del significado se nos vuelven invisibles, porque las palabras mismas parecen impregnadas con significado cuando las usamos todo el tiempo. De modo que es difícil incluso entender la pregunta "¿en virtud de qué cosa mi palabra 'rojo' significa un tipo particular de impresión sensorial?". Querríamos responder: "En virtud de que significa 'rojo' ". Pero esto sencillamente repite lo que queremos explicar, como si dijéramos: "En virtud de que significa *este tipo de impresión de color*".

Wittgenstein no ofrece un análisis del significado, pero sí trata de identificar ciertos intentos de usar el lenguaje que hacen que pierda sus amarras, de manera que la distinción entre la aplicación correcta e incorrecta del término en cuestión deja de existir, y todo lo que tenemos es la *palabra* (normalmente en bastardilla). El creía que el mítico lenguaje pri-

vado de experiencia, al estar completamente separado de nuestra conducta y de las circunstancias que pueden ser observadas por otros, y que se usa para expresar un escepticismo radical sobre la experiencia de los otros, es un ejemplo de esto. No podemos verlo, porque confiamos inconscientemente en nuestra comprensión de los conceptos experienciales del lenguaje público, cuyas condiciones de significado están ocultas en su aplicación subjetiva a nosotros mismos.

Si, en principio, nadie más que yo puede saber si estaba usando la palabra "rojo" (en tanto un término para un tipo de impresión visual) de la misma manera en dos ocasiones diferentes, entonces yo tampoco puedo hacerlo, y no habría ninguna distinción entre usarla del mismo modo o no. Para que esta distinción tenga sentido no es suficiente, según Wittgenstein, que usemos simplemente la palabra una y otra vez, diciendo: "Se trata de la diferencia entre usarla para una impresión que es realmente roja y usarla para una que no es roja". Naturalmente, *si* hay una diferencia, así es como la vamos a describir. Pero se debe decir algo más sobre hasta dónde podría llegar esa distinción.

El problema en estos casos es que tendemos a usar una imagen cuya aplicación habitual se encuentra bloqueada: por ejemplo, nos imaginamos dos cuadros de color, uno al lado del otro, que parecen diferentes. ¿Pero qué quiere decir esta imagen al ser aplicada al caso en que uso la misma palabra para dos impresiones de color separadas temporalmente, y no hay ninguna forma concebible de ningún tipo para compararlas, que no sea mi convicción, en particular, ninguna de las evidencias que usaríamos habitualmente para el dominio público: los objetos que estoy mirando, mi estado físico y la correlación entre mi identificación general de impresiones de color y la de otras personas? Plantear esta pregunta no es simplemente adoptar una posición verificacionista, ya que se podría mostrar que ciertas afirmaciones inverificables tienen sentido de este modo. Pero esto requiere que un tipo de imagen o expresión tenga alguna base de aplicación concebible si debemos creer que el sentido que nos da de comprenderlo no es ilusorio.

Que mi aplicación de una palabra en un caso determinado sea correcta depende de lo que yo quiera significar con ella. Pero lo que significo con ella es algo referido a mí mismo. De modo que, si tiene sentido la posibilidad de corregir lo que digo

de hecho con referencia a lo que debo decir a la luz de lo que significo, tenemos que encontrar una comparación entre mi uso convencido actual y algo más sobre mí mismo: sobre lo que estoy haciendo, en un sentido más amplio. Esta distancia crucial es lo que desaparece si la aplicación de mi lenguaje a mi experiencia es imaginada como lógicamente aislada de todo lo demás, como en el mítico lenguaje privado. No nos queda *nada más* que mis aplicaciones sucesivas de un término, sin nada para compararlas que pueda contar como una regla de su aplicación, es decir, sin el criterio que determina si la apliqué correctamente en cada caso.

El que su teoría satisfaga esta necesidad del lenguaje natural es un tema central en la interpretación de Wittgenstein, que fue sacado a luz recientemente por las sugerencias de Saul Kripke y las respuestas que recibió,[3] a las que se suma la de Pears. Pears explica el argumento wittgensteiniano de que la regla para la aplicación de un término no puede ser capturada por un "talismán mental instantáneo", ya que este no podría por sí mismo determinar la distinción entre correcto e incorrecto para una serie indefinida de futuras aplicaciones: dado que este talismán podría ser interpretado de más de una manera, deberíamos tener que explicar qué es lo que determina la corrección o no.

Todo lo que podemos hacer en esta dirección, según Wittgenstein, es describir posibles bases de corrección para el uso que hace un individuo de un término, que este reconoce como válidas. Según Pears, hay dos tipos principales de bases: la comparación con el uso de los demás y la calibración según objetos externos estándar que pueden ser reidentificados independientemente de la aplicación del término en cuestión. Al enfatizar la importancia de los objetos estándar, Pears se une a Colin McGinn.[4]

Ante la posibilidad de que esta base de corrección sea suficiente para suponer la existencia de una regla, podemos pensar que un lenguaje, que es privado, en el sentido de que fue inventado y utilizado exclusivamente por un ser solitario, no sería lógicamente imposible, pero tampoco sería *lógicamente* privado, ya que, en principio, otros podrían aprenderlo y, si lo hicieran, la comparación con el uso de los demás sería un control de si su inventor está empleando correctamente los términos del lenguaje.

Si un enfoque de esta clase es correcto, entonces vemos que apelar a la corrección o incorrección del uso particular de un término es simplemente más uso, por parte del usuario y otras personas, de ese y otros términos, y no algo de un tipo fundamentalmente diferente. Si superamos todos los desafíos que se nos presentan, y respondemos a cada desafío actual por medio de comparaciones que no dan origen a nuevos desafíos, entonces podemos considerar nuestra aplicación como correcta. Pero no hay ningún estándar platónico con el cual podríamos convalidar los juicios a los que apelamos en nuestra confirmación original. Más allá de un cierto punto, obedecemos "ciegamente" —como dice Wittgenstein— las reglas de nuestro lenguaje, y, si no lo hacemos, la idea del uso válido de un término sería imposible. Nuestro uso debe descansar, últimamente, si no en la primera instancia, en juicios que no podemos evitar hacer. Según Pears "mi obediencia es 'ciega', no porque dejo afuera consideraciones que pueden haberme influido... sino porque, cuando llego a los fundamentos, ... no hay más consideraciones, dudas o justificaciones. Ni siquiera tengo que escuchar la regla, porque habla a través de la aplicación que hago de ella".

Un tema sacado a luz por el análisis de Kripke es si esta posición —de que el único estándar de juicio es un círculo de juicio más amplio— es una forma de escepticismo. Al describir la posición de Wittgenstein como una "solución escéptica" al problema de las reglas, lógica y significado, Kripke no quiso expresar que Wittgenstein aceptaría tal encasillamiento. Como dice Pears, Wittgenstein pensó que el escepticismo surge de la necesidad de un estándar superior de corrección que es de un tipo imposible, platónico, en tanto que el estándar naturalista de lo que nosotros hacemos es todo lo que tenemos y necesitamos.

Pero, como piensa Pears, esto no resuelve la cuestión. Podemos llamar a alguien escéptico, aunque no acepte esa denominación, si es que pensamos que su versión de la base de una forma de pensamiento no proporciona una adecuada fundamentación a sus pretensiones, que van más allá de lo que esa base podría garantizar. Y podemos hacerlo, aunque no podamos esgrimir una fundamentación alternativa. Así, quien piensa que la ética sostiene cierto tipo de objetividad, dirá que el emotivismo es una teoría ética escéptica, aun si se presenta como el

análisis de los enunciados éticos ordinarios, aunque no pueda ofrecer una fundamentación alternativa de la objetividad de la ética. Lo mismo puede decirse del fenomenalismo como un análisis de los enunciados sobre el mundo externo.

La pregunta en este caso es si la teoría de Wittgenstein de que no hay ningún contenido en la idea de corrección en la aplicación de cualquier término universal que no sea su uso en un sentido amplio, debilita mucho la afirmación de objetividad que tenemos al entender esos términos y lo que creemos que estamos haciendo cuando pensamos con ellos. Creo que es así, que si Wittgenstein está en lo cierto, nuestro lenguaje no es lo que pensábamos. Lo que pensábamos era que, al seleccionarlos sobre la base de un uso compartido, llegaríamos a significar con "más" una determinada función aritmética, o con "rojo" un determinado color, y que este "significado", cualquiera que fuese y cualquiera que fuese nuestro uso de él, determinaba la diferencia entre la aplicación correcta e incorrecta del término, sobre la base de un abanico de posibles casos infinitamente más allá de todos los usos del término, y lo determinaba de un modo independiente de todo uso futuro. En otras palabras, el significado alcanza mucho más allá de las condiciones de aplicación concebibles, y los criterios limitados con los que controlamos su uso nos dan sólo evidencia de significado.

La respuesta de Wittgenstein es que esta idea es sólo aparentemente inteligible, y que su posición no desmantela nada posible y, menos aún, nada real. Pero incluso si él está en lo cierto, el que su solución sea escéptica depende de si esta ilusión de un sentido de largo alcance es interna a nuestro lenguaje, o si resulta de imponerle un artificio filosófico que es un estándar irreal externo. Podemos preguntarnos qué evidencia hay de que nuestro lenguaje hace afirmaciones más fuertes de las que puede sostener. Esto trae aparejada la pregunta de si la aparente naturalidad de los problemas escépticos en la filosofía misma aportan dicha evidencia. La falta de satisfacción y la sensación de presión que producen a menudo los argumentos de Wittgenstein sobre quienes los encuentran apremiantes aportan cierta evidencia a favor de la visión pesimista. Pero Pears, usando una imagen llamativa, dice: son como la gente a la que, después de haber estado hipnotizada, se le dice que no están pisando suelo firme sino un angosto puente sobre un desfiladero, por lo que la única manera de hacerlos caminar es decirles que hay parapetos a cada lado.

Al final del libro, Pears expresa cierta simpatía hacia la idea de que, si aplicamos la versión de Wittgenstein de las reglas a la lógica y a la matemática, resulta paradójica. No veo ninguna diferencia entre el caso de la lógica y la experiencia. Que la aplicabilidad supuestamente infinita de un término descriptivo debe descansar sólo en el acuerdo finito en su uso, es tan paradójico como que la expansión infinita de una serie matemática dependa de algo semejante. Nuestra creencia en la realidad concluyente de la expansión infinita o de la secuencia sin fin de casos no parece sostenerse en los fenómenos naturales humanos que Wittgenstein arguye que tenemos todos nosotros. Pears nota el paralelismo entre ambos casos, pero me parece que se equivoca al pensar que esta dificultad será más fácil de digerir en el caso de la filosofía de la mente, la filosofía general y del lenguaje que en el caso de la posición comparable en la filosofía de la matemática.

Sin embargo, el libro representa un gran logro, poniendo el énfasis en la continuidad de las preocupaciones filosóficas más profundas a lo largo de la vida de Wittgenstein. Sobre todo, la presencia de las motivaciones filosóficas tradicionales. Esta gran figura nos ocupará por un buen tiempo, y el estudio de Pears representa una contribución muy importante para el esfuerzo de lograr un dominio de sus ideas y de lo que las motivó.

Notas

1. En tanto que el *Tractatus Logico-Philosophicus* era un tratado genuino, los escritos posteriores de Wittgenstein eran más una versión destilada de la actividad extraña e inconexa del pensamiento y de la discusión filosófica en sí misma, lo cual, en la mayoría de los escritores, resulta en un producto altamente estructurado y completamente diferente a los procesos que lo produjeron.

2. *Philosophical Review*, julio de 1968. Pears enfatiza correctamente la importancia de estas notas, escritas por Wittgenstein en inglés entre 1934 y 1936, que parecen contener la primera versión del Argumento del Lenguaje Privado.

3. Saul Kripke, *Wittgenstein on Rules and Private Language* (Harvard University Press, 1980).

4. Colin McGinn, *Wittgenstein on Meaning* (Blackwell, 1984).

4

Chomsky: Lingüística y teoría del conocimiento*

Este ensayo fue elaborado a partir del trabajo que presenté para el volumen del Simposio organizado por el Instituto de Filosofía de la New York University los días 12 y 13 de abril de 1968. Yo había estado preocupado por las ideas chomskianas sobre la filosofía de la psicología desde los años 1960 y, cuando llegué a Harvard como estudiante de posgrado, encontré que los cursos más interesantes del área de Boston eran los de Chomsky en el MIT.

Ahora tengo mis dudas respecto del final del artículo, ya que parece sugerir que un conocimiento a priori no puede depender de las contingencias de nuestra constitución mental, lo cual implica claramente una confusión entre dos clases de "dependencia". Del hecho de que la posibilidad de justificación depende de ciertas contingencias, como el acuerdo intersubjetivo, no se sigue que esas contingencias formen parte de la justificación. Pueden ser meramente condiciones de comprensión, donde lo que se entiende no es contingente, sino necesario y, una vez que es comprendido, autoevidente. Nuestra habilidad para la lógica es contingente respecto de muchas condiciones, físicas y mentales, que no hacen que la lógica sea menos a priori. Sin embargo, esto no afecta para nada el caso de la gramática: creo todavía que la capacidad innata de aprendizaje del lenguaje no es una forma de conocimiento innato, porque no puede ser traído a la conciencia, y porque la corrección de sus resultados requieren confirmación empírica.

* Reimpreso con autorización de *Language and Philosophy*, Sidney Book, editor (New York University Press, 1969).

Hay alguna razón para creer que los aportes de Chomsky sobre la contribución innata a la adquisición del lenguaje se apoyan en temas cognoscitivos: en disputas sobre la existencia de un conocimiento a priori, por ejemplo. Si Chomsky tiene razón, la gramática representa un ejemplo contundente de las fuertes restricciones innatas sobre la forma del pensamiento humano y un objeto natural de fascinación filosófica.

No voy a analizar aquí la corrección de la posición de Chomsky respecto de la importancia y tamaño de dicha contribución innata, ni la adecuación de su fundamentación. El objetivo de este artículo es investigar cuáles son las consecuencias gnoseológicas de las hipótesis empíricas chomskianas sobre el aprendizaje del lenguaje, si es que son *correctas*. Dividiré el análisis en dos partes: primero, veré cómo pueden formularse las hipótesis de Chomsky del modo más adecuado posible y, específicamente, cómo puede calzar el concepto de conocimiento en esa formulación. En segundo lugar, consideraré el apoyo de sus hipótesis en el estado epistemológico de nuestro conocimiento de los lenguajes naturales, y también sus implicaciones respecto de otros tipos de conocimiento, algunos de los cuales son, a veces, considerados a priori.

I

El siguiente fragmento, de la página 58 de *Aspects of the Theory of Syntax,*[1] ofrece un enunciado claro y breve de la posición de Chomsky:

> Parece claro que la adquisición del lenguaje se basa en el descubrimiento del niño de algo que es, desde el punto de vista formal, una teoría profunda y abstracta —una gramática generativa del lenguaje— muchos de cuyos conceptos y principios se relacionan con la experiencia en forma remota por medio de largas e intrincadas cadenas de pasos inconscientes casi inferenciales. Un análisis del carácter de la gramática que está siendo adquirida, la calidad deteriorada y la extensión poco amplia de los datos disponibles, la llamativa uniformidad de las gramáticas resultantes y su independencia de la inteligencia, motivación y estado emocional, en un gran espectro de variación, nos dan poca esperanza de que un organismo inicialmente no informado en cuanto a su carácter general pueda aprender gran parte de la estructura del lenguaje.

Creo que Chomsky quiere decir que tenemos aquí un caso genuino de conocimiento innato. Sus amables referencias a los racionalistas sugieren que sí. Más aún: la alternativa a un organismo inicialmente *no* informado respecto del carácter general de la estructura de los lenguajes naturales parecería ser un organismo inicialmente *informado* respecto de ese carácter general. Y, en otra parte (p.27), Chomsky habla de atribuirle al niño un conocimiento tácito de los universales lingüísticos. En todo caso, los argumentos de Chomsky sobre la adquisición del lenguaje sugieren a los estudiosos de gnoseología, así como me lo sugieren a mí, que nos encontramos aquí con un ejemplo de conocimiento innato. Es esta interpretación filosófica natural la que quiero examinar, de modo que el resto de este ensayo no se referirá a las posiciones filosóficas chomskianas, sino sólo a las implicaciones filosóficas de sus concepciones lingüísticas.

La primera pregunta es, entonces, si es correcto llamar conocimiento a la contribución inicial del organismo para el aprendizaje del lenguaje tal como sostiene Chomsky. Comencemos con un argumento natural aunque malo, a favor de una respuesta negativa a la pregunta. El argumento adopta la forma de una reducción.

¿Por qué la decisión de aplicar el concepto de conocimiento no nos obliga a adjudicarle conocimiento innato, o incluso conocimiento a priori, al sistema digestivo humano (o, quizás más bien, al ser humano en virtud de la conducta de su aparato digestivo)? Ya que, sin haber sido entrenado, enseñado o condicionado, el individuo puede ajustar el medio químico en su estómago para desmenuzar los alimentos digeribles que se le introducen, y rechazar —a veces, por la fuerza— lo que no es digerible. Esta formidable tarea de clasificación y respuesta variable es llevada a cabo incluso por niños pequeños, de modo que no puede ser aprendida solamente con la experiencia.

Es claro que el niño pequeño no es consciente de los principios que gobiernan sus secreciones gástricas, ni tampoco lo es un adulto, salvo que haya estudiado fisiología. Pero ni los niños ni los adultos que estudiaron lingüística son conscientes de las reglas gramaticales de su lenguaje o de los principios por los que adquieren la habilidad de hablar el lenguaje gobernado por dichas reglas, sobre la base de su exposición a una subclase de oraciones de ese lenguaje. Las razones que podemos ofrecer a favor de un conocimiento innato del carácter ge-

neral de la estructura lingüística cuentan también a favor de la posición de que hay un conocimiento innato de los medios químicos adecuados para digerir diferentes clases de alimentos. O bien ambos son ejemplos de conocimiento innato, o bien ninguno de los dos lo es. La segunda posibilidad es la más plausible. Podríamos decir que en ambos casos hay una *capacidad* innata extremadamente importante —para discriminar alimentos y digerirlos, o para adquirir el dominio de lenguajes naturales que tienen cierto tipo de estructura—, pero no la llamaríamos, en ninguno de los dos casos, *conocimiento* innato.

El problema de este argumento es que pasa por alto la diferencia entre las operaciones que somos capaces de realizar en cada uno de los dos casos. En el caso de la digestión, la operación no es una acción. Nuestro organismo no tiene tampoco que ser consciente de los datos sobre los que se basa dicha operación, es decir, la variedad de alimentos que se introducen en el estómago. Por el otro lado, en el caso de aprendizaje del lenguaje, es fundamental la aprehensión consciente de los datos (aunque sean limitados), y lo que el individuo puede hacer como resultado de su capacidad lingüística es hablar y comprender oraciones.

Más aún, el ejercicio de la capacidad implica *creencias*: por ejemplo, que cierta combinación de palabras es o no una oración de una determinada lengua. Por el contrario, de alguien que vomita una ostra en mal estado nunca diríamos que cree que no es digerible. Aunque no tengamos un análisis adecuado de la distinción, es claro que ciertos métodos de respuesta y discriminación garantizan la atribución de creencias y actitudes, en tanto que otros no. Sólo respecto de los de la primera categoría será adecuado analizar si son evidencia de conocimiento o no.

Es claro que estos conceptos cognitivos sólo pueden ser aplicados a la capacidad y acción lingüísticas en ocasiones particulares. La cuestión es si los conceptos de conocimiento y creencia pueden aplicarse a niveles más altos de generalidad y abstracción en la descripción de la capacidad lingüística del individuo y, en última instancia, en la descripción de su capacidad de adquirir esa capacidad.

Podemos distinguir dos tesis: (1) que la capacidad general para producir un conjunto de acciones, cada una de las cuales proporciona una instancia de conocimiento, es en sí misma una

instancia de un conocimiento más general; (2) la capacidad general para adquirir otras capacidades, cada una de las cuales es una instancia de conocimiento es, en sí misma, una instancia de un conocimiento todavía más general. La primera tesis es más plausible que la segunda, pero ambas son necesarias para posibilitar el ascenso inferencial desde los casos de conocimiento lingüístico revelado en oraciones particulares a los universales lingüísticos de los cuales se supone que depende el aprendizaje del lenguaje.

Vayamos paso por paso: Parece obvio que podemos hablar de un conocimiento lingüístico cuyo objeto no sea meramente la gramaticalidad o el significado de una oración particular, sino algo más general. (En realidad, es dudoso que podamos hablar de conocimiento en un caso particular, salvo que podamos hablar de conocimiento en un nivel más general.) Por ejemplo, podemos atribuirle al hablante común de inglés, sobre la base de innumerables acciones y respuestas particulares, el conocimiento de que la forma plural de un sustantivo se forma normalmente agregando una "s", y que, entre las excepciones a esta regla, está la palabra "man", cuyo plural es "men". Ahora *podríamos* verificar esta atribución si el individuo logra de hecho enunciar la regla, pero alguien puede poseer un conocimiento general de la regla aunque no sea capaz de enunciarla. Puede ser que nunca haya escuchado las palabras "plural" y "sustantivo", por ejemplo, o puede ser incapaz de formular el principio de cualquier otra manera. Esta es la situación habitual cuando acudimos a los principios gramaticales más complicados del habla inglesa. Sólo los gramáticos profesionales pueden enunciar esas reglas, y a veces ni siquiera ellos.

¿Bajo qué condiciones podemos describir el conocimiento de un lenguaje gobernado por ciertas reglas como el conocimiento de esas reglas? Tomemos otro tipo de conocimiento que no puede ser explicado por sus usuarios, a saber, el conocimiento inconsciente, en el sentido psicoanalítico común. Este es, naturalmente, un fenómeno diferente al del conocimiento de las reglas gramaticales, pero tiene un rasgo importante que, tal como me señaló Saul Kripke, puede aplicarse al caso de la lingüística. La atribución psicoanalítica de conocimiento o creencia inconsciente —o motivos inconscientes— no depende simplemente de la posibilidad de organizar las acciones y respuestas del sujeto de acuerdo con el supuesto material inconsciente. Más

aún, aunque el sujeto no formule su conocimiento o actitud inconsciente espontáneamente, y si se le pregunta, a veces lo niega, es por lo común posible conducirlo a que *vea* que el enunciado en cuestión expresa algo que sabe o siente. Eventualmente, si se lo presentamos en forma suficientemente clara y en las circunstancias adecuadas, puede reconocer el enunciado como una expresión de su propia creencia. Por lo tanto, se puede traer a la conciencia al menos una parte de lo que era inconsciente. Es fundamental que su reconocimiento *no* se base meramente en la observación de sus propias respuestas y conducta, y que llegue a reconocer la corrección de la atribución desde dentro.

Me parece que tenemos una buena razón para hablar de conocimiento y creencia cuando un reconocimiento de este tipo es, en principio, posible, aun en casos en que los principios o enunciados relevantes no son todavía reconocidos conscientemente, o incluso en casos donde nunca serán explícitamente formulados. Aunque el conocimiento de las reglas del lenguaje es en otras formas bastante diferente al conocimiento inconsciente revelado por el psicoanálisis, las formulaciones certeras de las reglas gramaticales evocan a menudo el mismo sentido de reconocimiento por parte de los hablantes que se sometieron a ellas durante años que se evoca con la formulación explícita del material reprimido que influyó su conducta durante años. Puede ocurrir, incluso si los principios gramaticales se formulan en un lenguaje técnico que posiblemente requiera una dosis de esfuerzo para poder ser dominado.

En tanto *sea* posible conducir con esfuerzo al hablante a un reconocimiento genuino de una regla gramatical como expresión de su comprensión del lenguaje, más bien que como una mera creencia basada en la observación de casos, de que la regla describe de hecho su competencia, es aceptable, en mi opinión, atribuirle al hablante el conocimiento de esa regla. Es aceptable, aunque posiblemente nunca pueda ser presentado con una formulación de la regla y, consecuentemente, el sujeto nunca pueda llegar a reconocerla conscientemente.

Pero si no se satisface la condición del reconocimiento, la atribución de conocimiento y creencia me parece más dudosa. Y esto hace dudosa la posibilidad de llevar a cabo la atribución de conocimiento a cualquier nivel de generalidad o abstracción más elevado que el que está implicado en la especificación de

reglas gramaticales para un lenguaje natural particular. Incluso algunas de esas reglas son extremadamente abstractas. Pero cuando consideramos la supuesta contribución innata al aprendizaje del lenguaje, pasamos a un nivel bastante distinto; es dudoso que los principios de ese dispositivo de adquisición lingüística, cuando son formulados, puedan evocar un reconocimiento interno en los individuos que actuaron de acuerdo con ellos.

Las reglas de una gramática particular tratan en parte con expresiones reconocibles y, en su formulación, conservan alguna conexión con la experiencia consciente que tiene el hablante de su lenguaje. Esta conexión, en el caso de los universales lingüísticos de la clase que Chomsky sugiere que están presentes en forma innata, es más remota. Un ejemplo que ofrece Chomsky es la propuesta de que el componente sintáctico de una gramática debe contener reglas transformacionales. Esta condición altamente abstracta se supone que se aplica a *todas* las lenguas, y que determina la forma en que un niño adquiere conocimiento de la gramática de su lengua materna al ser expuesto a segmentos del habla. ¿Pero se supone que el hablante podría, llegar a reconocer ese principio como la expresión de una suposición que hizo en aquel momento (una vez que el principio había sido formulado y que había captado su significado)? Esto podría ser una posibilidad, pero las condiciones de adecuación explicativa que Chomsky acepta no me parecen exigirla. La adecuación explicativa es, en sí misma, naturalmente, un requisito muy fuerte. Pero se puede mostrar que una hipótesis la satisface sobre la base de la observación externa del proceso de aprendizaje de la lengua. El test adicional de preguntarle al que aprende la lengua si puede reconocer el principio como aquel que lo estaba activando durante el proceso parece irrelevante. No parece que se requiera que ese reconocimiento interno sea *alguna vez* posible o accesible, independientemente de cuánto esfuerzo ponga en ello.

Por ello no me gusta extender el concepto de conocimiento, así como los conceptos relacionados de creencia y suposición, a la descripción de esa capacidad innata que le permite al niño adquirir conocimiento de un lenguaje —cualquier lenguaje natural— sobre la base de datos más bien exigua. No toda capacidad innata para adquirir conocimiento es necesariamente, en sí misma, una instancia de conocimiento, aunque su estructura sea bastante compleja.

II

Los problemas planteados hasta aquí se refieren realmente a la atribución de *creencias* o *suposiciones* sobre la base de la habilidad de aprendizaje del lenguaje. Pero la cuestión que es más interesante desde el punto de vista del conocimiento es si la otra condición fundamental de conocimiento —a saber, la condición de justificación— podría ser satisfecha en estos casos. No hay acuerdo respecto de la naturaleza de esta condición, pero sería posible analizar el problema sin internarnos en ese laberinto.

El problema es que podemos imaginar que casi todas las creencias están presentes en forma innata, o bien que hay una tendencia natural a desarrollar esa creencia, como resultado de ciertas experiencias mínimas, pero esto no es suficiente para atribuir conocimiento. No cualquier creencia a la que inevitablemente llegamos está justificada *ipso facto,* aunque sea verdadera.

Aun si alguien descubre que puede especificar la raíz cuadrada de un número entero de cuatro puntos decimales sin reflexión o cálculo, el carácter innato de su habilidad no garantizaría la validez de sus respuestas. La fundamentación de su conocimiento de la raíz cuadrada sería más bien que él u otra persona podrían verificar, por medio del cálculo, caso por caso, si el número que, sin reflexionar, era considerado la raíz cuadrada de un número entero dado era, de hecho, la raíz cuadrada. Gracias a esta evidencia, su creencia en un caso dado podría ser considerada como una fuerte evidencia de su verdad.

El fenómeno del aprendizaje del lenguaje que Chomsky describe es diferente porque revela una capacidad innata que todos compartimos. Todos los hablantes de inglés alcanzan un acuerdo con respecto a la obediencia a ciertas reglas gramaticales, y lo adquieren naturalmente y sin cálculo después de una exposición a la lengua. Ahora bien, la propensión innata de un solo individuo para alcanzar esas reglas no garantiza por sí misma que esas reglas sean las reglas de la lengua que él habla. Esto depende de una conformidad más general a las mismas reglas, por parte de todos los hablantes de la lengua, que está garantizada por la universalidad de esas mismas propensiones innatas. Por lo cual, si cualquier individuo dado sabe

que sus intuiciones lingüísticas sobre oraciones que nunca antes había encontrado, así como sus propias producciones lingüísticas originales, están de acuerdo con las intuiciones lingüísticas de otros hablantes de su lengua, entonces puede considerar que sus tendencias innatas le otorgan una fuerte evidencia a favor de su adecuación. No digo que nosotros de hecho nos alejamos de nuestras intuiciones lingüísticas, para validarlas de este modo. Lo que estoy sugiriendo es sólo que, gracias a una justificación de este tipo, podemos describir en forma plausible lo que nos dan nuestras propensiones lingüísticas innatas en tanto *conocimiento* de la lengua.

En ningún caso, el solo hecho de que nuestra tendencia a llegar a cierta creencia sea innata nos habilita a hablar de un caso de conocimiento. En el caso especial del lenguaje, en el que las reglas son, simplemente, las que siguen los hablantes competentes, una tendencia innata universal para llegar a ciertas reglas es suficiente para garantizar su certeza; pero cualquier individuo debe siempre saber que él está en conformidad con la tendencia universal, a fin de saber si sus intuiciones lingüísticas son correctas. Esta es una cuestión empírica. En cualquier caso individual, la supuesta contribución innata al aprendizaje de la lengua puede ser evaluada respecto de su certeza como una fuente de conocimiento del lenguaje. Puede ser que nadie adopte nunca este tipo de evaluación y que la tendencia innata a construir la gramática de la lengua de uno también incluya una tendencia innata a suponer que los otros hablantes también la construirán del mismo modo. En efecto, parece probable que así sea. Pero esa suposición también está abierta a otro tipo de evaluación con otros medios.

De modo que el factor innato que, según Chomsky, tiene que subyacer a nuestra capacidad de aprendizaje del lenguaje no tiene parecido alguno con el tipo de fundamentación incuestionable, cognoscitivamente irrebatible, sobre la que algunos filósofos trataron de fundamentar el conocimiento humano, y que normalmente se llama conocimiento a priori o innato. Lo que se buscaba bajo este nombre es algo que no está en sí mismo abierto a evaluaciones y dudas de conocimiento, algo cuyo opuesto es inimaginable.

Lo que Chomsky ofrece es un sistema de propensiones innatas, con las que nos encontramos convenientemente dotados. Es perfectamente posible imaginarnos constituidos de otra

manera, pero no lo estamos. Una mera tendencia innata a creer ciertas cosas o a actuar de cierto modo, no importa cuán universal sea, no es un conocimiento a priori. Incluso Hume pensó que todos compartimos una propensión natural a creer que el sol saldrá mañana. Pero señalar el fenómeno natural del acuerdo humano, determinado en forma innata, es simplemente deshacerse de la exigencia gnoseológica que origina la búsqueda del conocimiento a priori.

De hecho, una movida así está relacionada muy de cerca con la posición de Wittgenstein. La principal diferencia es que Wittgenstein la aplica en una forma mucho más general, y no sólo al aprendizaje de la lengua. El sostiene que si alguien sigue cualquier cadena de justificación cognoscitiva lo suficientemente lejos, se llega al final a un fenómeno de acuerdo humano —no un acuerdo convencional, sino un acuerdo natural, determinado en forma innata— del cual depende la aceptación de esa justificación. El supone que esto ocurre tanto en la justificación empírica o deductiva. Si tiene razón, los procedimientos por los cuales sometemos una contribución innata a evaluación cognoscitiva dependerán, ellos mismos, de otra contribución innata. Y si, en cada etapa, lo que alcanzamos es sólo un rasgo contingente de nuestra constitución, entonces no hay ninguna fundamentación incuestionable a priori sobre la que se apoye nuestro conocimiento, que depende de una red de respuestas y propensiones innatas, que simplemente nos son dadas.

Si esto es así, entonces la teoría del conocimiento es esencialmente imposible. En tanto los argumentos de Chomsky sobre el lenguaje sugieren que otras contribuciones innatas similares subyacen en otros fenómenos cognitivos, sugieren también que todo el conocimiento se encuentra en dificultades similares: carece de una última fundamentación. A veces, en el caso del lenguaje, podemos justificar nuestra confianza por medio del producto de nuestro mecanismo innato. Cualquier hablante que use exitosamente el lenguaje para comunicarse con los demás puede emplear esta clase de evidencia. Pero la aceptación de esta evidencia puede, a su vez, depender de principios innatos que, sin garantizar su propia justificación, forman parte de nuestra constitución básica. De manera que la tarea de justificación puede ser imposible de completar.

A pesar de que esto es incómodo desde el punto de vista del conocimiento, tiene compensaciones prácticas. Si tuviéramos que aprender con ensayo y error, o con ejercicios, a digerir la comida, nuestra supervivencia sería mucho más difícil. Pero afortunadamente, no tenemos que *saber* cómo digerir la comida, porque lo hacemos automáticamente. El aprendizaje del lenguaje puede ser similar. No necesitamos *saber* cómo construir la gramática de un lenguaje natural sobre la base de nuestra exposición a segmentos de la lengua en nuestra temprana infancia. Simplemente *llegamos* a dominar la lengua después de un período de exposición, y estamos convencidos de que los otros hablantes siguen las mismas reglas.

Puede ser cierto que en muchas áreas de actividad y experiencia humana, si tuviéramos que confiar en lo que podríamos saber sirviéndonos de otros medios empíricos o racionales, seríamos incapaces de sobrevivir. Pero si en estas áreas somos lo suficientemente afortunados como para tener una dote natural que nos prepara para enfrentar al mundo, no necesitamos un conocimiento que sería muy difícil, o quizás imposible, de alcanzar. En cambio, podemos dejarnos guiar por nuestras ideas innatas.

Notas

1. M.I.T. Press, 1965. [versión cast. Barcelona, Gedisa, 1999.]

5

Fodor: Los límites del espacio interno*

Este ensayo fue presentado en un simposio de la American Philosophical Association que se realizó el 26 de diciembre de 1968. Era una respuesta al artículo de Jerry Fodor, "The Appeal to Tacit Knowledge in Psychological Explanation", publicado en The Journal of Philosophy 65 (1968), *p. 627. Estoy algo sorprendido por mi tono agresivo, pero todavía me parece que la imposibilidad de reducir la intencionalidad a cualquier tipo de regularidad sigue siendo un obstáculo para el uso de la simulación mediante el ordenador como la clave para la esencia de la mente. Aunque haya mucho que aprender en esos programas, se ha perdido algo, que es el aspecto más difícil del problema.*

Suceden muchas cosas en nuestro cuerpo y mente, a las que no tenemos acceso introspectivo. Con el entrenamiento adecuado, podemos extender el alcance de nuestra percepción no observacional e introspectiva. Sin embargo, somos organismos sumamente complicados, demasiado complicados como para examinar cada detalle de nuestro funcionamiento. Afortunadamente, una gran cantidad de lo que nosotros y nuestro cuerpo hacemos no requieren atención ni resultan de aprendizaje, o de un aprendizaje mínimo. Esto es verdad, no sólo de los procesos totalmente involuntarios, como curarse, la coagulación de la sangre, la espermatogénesis y la ovulación, sino también de las alteraciones en las células nerviosas y musculares que for-

* Reimpreso con autorización de *The Journal of Philosophy* 66, 1969.

man la subestructura de la acción intencional común, desde atarse los cordones hasta tocar pasajes *spiccato* en violoncello. Nos ocupamos de los kilogramos, pero no prestamos atención a los microgramos. Como el interesante artículo de Fodor señala, lo mismo es verdad para las operaciones mentales (cualquiera fuera su relación con la función cerebral). Nuestra percepción profunda depende de factores en los que no pensamos nunca, y la relación precisa entre otras experiencias y la percepción de profundidad o la estimación de su tamaño debe ser determinada con experimentos psicológicos. No es más accesible a la introspección que el mecanismo con el cual se forma el tejido cicatrizante.

Hechos de este tipo traen a luz una pregunta general. ¿De todas las cosas que hacen el cuerpo y la mente de un hombre, cuántas podemos decir que *él* hace? ¿Todas las respuestas y operaciones que su cuerpo, su cerebro y su mente llevan a cabo dócilmente, y sin las cuales no podría hacer nada, son cosas que es capaz de hacer o sabe cómo hacer? Si comenzamos con algo que cualquiera describiría como una acción, por ejemplo, atarse los cordones, podríamos desglosar la acción, parte por parte, más allá de los dedos, músculos y percepciones de superposición, hasta el nivel de los cambios en la permeabilidad de las paredes celulares y el gradiente potencial de las sinapsis nerviosas y, más allá aún hasta la alteración en las macromoléculas, en el núcleo de la célula o en los sucesos subatómicos de los cuales ello depende. En algún punto será claro para todos que, si vamos a la profundidad suficiente, hemos perdido a la persona, encontrándonos no con los medios con los que se ata los cordones, sino con la infraestructura fisiológica y mental de sus acciones. Si hay una línea entre la persona y el resto del sistema altamente organizado en el que su vida transcurre, ¿dónde está, y qué clase de línea es? Voy a discutir el artículo de Fodor a la luz de esta pregunta y de la respuesta que él le ha dado en forma implícita.

Dos cosas parecen claras sobre el tipo de caso que Fodor quiere explicar parcialmente en términos de conocimiento tácito. Primero, si lo entiendo correctamente, este tipo de caso debe incluir una ejecución o funcionamiento terminal, que es más o menos consciente y que, además, nos da la base para atribuirle al agente un conocimiento no tácito: por ejemplo, el conocimiento

de cómo atarse los cordones, o el conocimiento de las distancias relativas de las cosas en su campo visual. El conocimiento tácito en estos casos subyace a un conocimiento explícito o habilidad más ampliamente reconocido. Esto significa que el conocimiento tácito no es atribuible a un individuo en la explicación de las múltiples operaciones inconscientes e involuntarias que lleva a cabo su cuerpo, desde la regulación del ritmo cardíaco hasta el aumento de los leucocitos en la sangre.

En segundo lugar, aun cuando la operación es, claramente, del tipo adecuado, no puede ser descompuesta indefinidamente como para alcanzar el conocimiento tácito que la explica. Podemos descomponer la acción de atarse los cordones en los movimientos y combinaciones de átomos individuales, y, usando la imagen de Fodor, representar al individuo como un enorme ejército de pequeños hombrecitos, cada uno de ellos correspondiente a cada átomo, que obedecen las órdenes que les llegan a través de una compleja cadena de órdenes impartidas por un hombrecito que está más o menos a cargo. Sin embargo, supongo que Fodor no estaría inclinado a atribuirle conocimiento tácito al nivel de la estructura atómica, porque no podemos decir que movemos los átomos en nuestros cuerpos; estos movimientos no son operaciones que ejecutamos, por medio de las que hacemos cosas más elaboradas.

Fodor marca el límite interno de la acción y, por lo tanto, del conocimiento por medio del concepto de *operación elemental.* "Una operación elemental", dice Fodor, "es una operación que el sistema nervioso normal puede ejecutar, pero de la que no puede ejecutar una parte propiamente dicha. Hablando intuitivamente, las operaciones elementales son aquellas que no tienen ninguna estructura interna relevante teóricamente" (p. 629). Esta no es una definición clara. Por ejemplo, no es claro si la definición excluye alguna operación con una parte propiamente dicha que pueda ocurrir en tanto componente de una operación diferente, aunque ninguna de sus partes propiamente dichas puedan ocurrir solas. Supongamos que lo que quiere decir es la posibilidad de ejecutar aisladamente cualquiera de sus partes propiamente dichas, y pasemos por alto por el momento los problemas sobre qué quiere decir "ejecutar" aquí. Quisiera sostener que no hay ninguna razón para creer que las operaciones que no son elementales para un organismo consisten en secuencias o grupos de operaciones que son elementales para ese organismo.

En primer lugar, una operación puede no ser elemental porque alguna de sus partes propiamente dichas puede ser ejecutada separadamente, aunque lo que permanece cuando esa parte es eliminada no puede ser ejecutado separadamente. En segundo lugar, puede haber más de una manera de desglosar exhaustivamente en sus partes elementales una operación no elemental. En tal caso, ¿cómo podemos decidir de qué "manera" el agente ha ejecutado esa operación? En tercer lugar, aunque hubiera sólo un conjunto de operaciones elementales en la base de alguna acción, el modo como se originan los movimientos elementales como parte de una acción más amplia puede no tener nada que ver con el modo como se originan cuando ocurren aisladamente.

Veamos algunos hechos. En la parte redondeada del pulgar hay un músculo llamado *abductor pollicis brevis*, que desempeña un papel fundamental para atarse los cordones, para ejecutar los pasajes *spiccato* en cello, y para otras cuestiones. (Produce el movimiento por el cual uno presiona la barra espaciadora en la máquina de escribir.) Este músculo consiste en varios miles de fibras musculares, y está enervado por aproximadamente trescientas motoneuronas, cada una de las cuales controla un subconjunto particular de fibras. El sistema que comprende una motoneurona y las fibras musculares que enerva se llama *unidad motora*. La activación de estas unidades motoras en combinaciones sutiles nos permite dar la nota correcta en una sonata de Beethoven o la forma característica de la primera letra de nuestra firma.

Recientemente, el neuroanatomista J.V. Basmajian descubrió que los sujetos humanos podían ser entrenados para activar a voluntad unidades motoras individuales seleccionadas del abductor *pollicis brevis*.[1] Los sujetos pudieron producir la imagen adecuada en un osciloscopio programado para registrar la descarga de las unidades motoras. Más adelante, algunos de ellos fueron capaces de hacerlo sin mirar al osciloscopio. Pudieron activar unidades motoras adyacentes separadamente, e incluso marcar ritmos de batería rápidos y elaborados, aunque comúnmente esas unidades no pueden descargarse de modo aislado. Cuando se les preguntó cómo podían hacerlo, no supieron contestar, así como no sabemos cómo nos las arreglamos para mover nuestros pulgares.

Recientemente, se avanzó aun más en este terreno. Fox y Rudell, dos psicólogos de la Universidad de Iowa, usaron técni-

cas de refuerzo estándar en el entrenamiento de gatos para que produjeran reacciones cerebrales específicas en respuesta a una señal visual.[2] Se elige una respuesta cortical arbitraria para el refuerzo: esta puede detectarse con electrodos implantados y se supone que probablemente no ocurrirá antes del entrenamiento. Se la pone bajo control operante: es decir, el gato aprende a usarla para conseguir alimento cuando tiene hambre. Los científicos no pudieron poner en correlación ninguna conducta notoria con la producción del suceso neural por parte del gato. Que yo sepa, este experimento no ha sido llevado a cabo con humanos.

Quise introducir estos ejemplos para mostrar que algunos componentes extremadamente minúsculos de la estructura física de un acto pueden, por sí mismos, alcanzar el estatus de acción intencional. Pero incluso si después de meses de entrenamiento yo fuera capaz de adquirir control individual de cada una de las unidades motoras del abductor *pollicis brevis*, esto no implicaría que cuando presiono la barra espaciadora de mi máquina de escribir estoy *activando* las trescientas unidades motoras a la vez. No hay duda de que cuando presiono la barra espaciadora, todas o la mayoría de las unidades motoras se activan a la vez. Yo no podría incluso ni siquiera enfrentar el desafío o, si lo intento, podría ser que nada sucediera, o podría producir un calambre terrible en mi pulgar. Por cierto, es concebible pensar que yo tendría éxito. Sin embargo, ¿qué razón hay para pensar que el método por el que se produce el movimiento común del pulgar tiene alguna relación interesante con el método por el que el mismo se produce por medio de la sumatoria de trescientas enervaciones motoras intencionales simultáneas? Por lo que entiendo, el peso de la información neurológica habitual no nos da la base para una visión atomista de la conducta de este tipo. Más aún, hay una gran cantidad de evidencia a favor de lo contrario, evidencia de que el cerebro (de alguna manera todavía no formulada) codifica los patrones de la conducta general compleja, en vez de recopilar los movimientos de los componentes. (Algo análogo parece verdad respecto del reconocimiento perceptivo de universales.) No es *necesariamente* falso que nos atamos los cordones o tocamos instrumentos musicales o hablamos por la composición, consciente o inconsciente, de movimientos nítidos de los dedos y de la lengua (sin mencionar las enervaciones individuales de las

unidades motoras). Simplemente, es falso. Una respuesta a la pregunta "¿Qué ocurre cuando decimos 'Beethoven'?" no es *ipso facto* una respuesta a "¿Cómo se dice 'Beethoven'?". Si lo entiendo bien, el modelo de Fodor del hombrecito y su batallón de subalternos no representa una dirección promisoria para la teoría neurológica actual.

Para volver a la cuestión del conocimiento, me parece que no tiene sentido aplicarle ese término a la capacidad integradora que permite la producción de una complicada secuencia de movimientos, cada uno de los cuales podría ser producido separadamente, pero que de hecho se producen de un modo diferente cuando ocurren como una parte de la secuencia. Esto es aun más evidente si la secuencia *no puede* producirse adecuadamente a través de una nítida producción secuencial de los movimientos componentes individuales. El conocimiento no tiene que ser expresable verbalmente, pero tampoco tiene que ser ejercitado conscientemente ni ser capaz de alcanzar conciencia con la reflexión adecuada.[3]

Sé que la expresión "conocimiento tácito" es un barbarismo técnico, y que hay algo peculiar en hacer que las intuiciones lingüísticas comunes se refieran a la cuestión de su aplicación. Sin embargo, me parece que hay tanto sentido en objetar esta degradación terminológica como en objetar el nombre "Salisbury steak". Además, si aplicamos el término "conocimiento", aunque sea calificado, a la capacidad que Fodor llama "conocimiento tácito", entonces agregamos un problema más a la oscuridad de una importante cuestión filosófica, que ya es de por sí oscura: a saber, dónde se deben trazar los límites internos del yo.

Puede ser totalmente imposible resolver este problema, pero, si lo intentáramos, el concepto de conocimiento sería uno de nuestros recursos importantes y, por eso, sería deseable alcanzar un acuerdo sobre ese concepto dentro de su hábitat natural, antes de extenderlo a un ámbito de fenómenos más amplio. El concepto de acción intencional es otro que debería ser examinado en este tipo de investigación. Atribuimos tanto conocimiento como acción a las personas individuales, los que excluyen muchas cosas que el organismo puede hacer pero que no les atribuimos a las personas. Ahora bien, puede ser que estos conceptos y las distinciones que traen a la luz no sean teóricamente interesantes. Puede ser (aunque lo dudo) que la idea de una persona, a la que estos conceptos están unidos, sea una

noción desfalleciente, y que probablemente no sobreviva los avances de la psicología científica y la neurofisiología. Quizá tengamos que echar mano de la idea de un organismo, o de un sistema orgánico. Pero no podemos saberlo sin antes haber tratado de entender el otro concepto.

Fodor, naturalmente, no querría abandonar el concepto de persona. El quiere extender el campo de los procesos y capacidades orgánicas que podemos atribuirle a la persona como sujeto, mediante la extensión de los conceptos de conocimiento, acción y juicio. Pero al darle al homúnculo demasiado para hacer, es posible que haya oscurecido el carácter especial de lo que, en realidad, podemos decir que hacen los seres humanos. Si yo supiera cuál es este carácter, se lo diría, pero no lo sé. Tampoco lo sabe Fodor, pero de todos modos usa esta idea, al apelar solamente al conocimiento tácito para explicar las *ejecuciones y las operaciones* humanas, y no sólo para explicar alguna función orgánica, incluso una altamente integrada como el control metabólico, que sería presumiblemente tan aplicable a la simulación computacional como el atarse los cordones. Pero, si admite la relevancia de una noción como ejecución en el nivel terminal, ¿por qué no podría admitirla en el nivel de la integración, donde se posibilita la ejecución compleja? En otras palabras, si la integración de los movimientos de atarse los cordones no es en sí una ejecución consciente o intencional, ¿por qué no le negamos el término "conocimiento" a la capacidad de producir esa integración, así como se lo negaríamos a la capacidad de atar cordones si el atarse los cordones no fuera en sí una ejecución consciente o intencional?

Permítaseme observar antes de terminar que no entiendo cuál es el propósito de Fodor al introducir una máquina en el argumento. Deberíamos averiguar cómo se produce la conducta del organismo, a fin de saber si una simulación computacional es óptima; si lo supiéramos, tendríamos todo el apoyo necesario con que podríamos contar para la descripción del individuo, en tanto aplicando conocimiento tácito, o siguiendo reglas. No puedo imaginarme qué agrega la simulación computacional, incluso suponiendo que la noción tenga sentido aunque la etiología del *output* de la máquina pueda ser representada con una secuencia de oraciones en inglés. Aunque es alarmantemente fácil tener la impresión contraria a partir de este artículo, su-

pongo que Fodor no sostiene que, dado que evidentemente la máquina tiene un conocimiento tácito de las reglas con las que opera, podemos inferir legítimamente, a partir de efectos similares a causas similares, que la persona tiene conocimiento tácito de las mismas reglas. ¿Pero entonces, qué *podemos* concluir del caso de este ordenador imaginario? Se puede pensar que por lo menos podemos inferir a partir del carácter computacional de la máquina el carácter computacional del organismo. Pero, dado que una analogía cercana en cuanto a esto es ya una condición de simulación óptima, no comprendo el objetivo de todo el procedimiento.[4]

Notas

1. *Science* 141 (1963) pp. 440-441. Véase también C.G. Phillips, "Changing Concepts of the Precentral Motor Area", en J.C. Eccles, comp., *Brain and Conscious Experience* (Springer, 1966), p. 40.

2. *Science* 162 (1968), pp. 1299-1302.

3. He desarrollado este argumento en "Chomsky: Lingüística y Epistemología", en el cap. 4 de este volumen. Permítaseme observar que una referencia a la conciencia o, quizás, a la intención en las condiciones para atribuirle conocimiento a una persona, meramente plantea el problema, sin solucionarlo. Debemos explicar precisamente estos conceptos, para entender qué quiere decir atribuirle una capacidad a una persona antes que sólo a su cuerpo o cerebro.

4. Sin embargo, en su análisis, Fodor señaló que la *evidencia* de una simulación por ordenador no necesariamente son producto de comparar sólo las formas en que se producen la conducta y su simulación. Sobre la base de otra evidencia, por lo tanto, se podría concluir que la simulación fue óptima y, de ahí, inferir que la conducta y el *output* de la máquina tuvieron la misma etiología.

6

Armstrong: Sobre la mente*

Este ensayo fue una reseña del libro A Materialist Theory of the Mind *de David M. Armstrong (Routledge and Kegan Paul / Humanities Press, 1968). Además de las esperadas críticas a Armstrong por su fracaso en tratar los problemas planteados por la subjetividad de lo mental, este ensayo contiene lo que era entonces un avance de las ideas de Kripke sobre identidad y necesidad, que yo conocía a partir de lecturas y conversaciones. Me llama la atención el hecho de que yo expresara algo de simpatía por el materialismo eliminativo, una posición que hoy considero ridícula.*

I

Gran parte de la literatura materialista reciente defiende un materialismo de sentido común: la concepción de que la identificación de lo mental y lo físico no requiere ningún ajuste o alteración de nuestros conceptos psicológicos corrientes. Se sostiene que el materialismo no es paradójico y no exige que abandonemos la concepción común del hombre. El libro voluminoso, lúcido y fácil de leer de Armstrong lleva la defensa de esta posición más lejos de lo que se haya llevado antes. Esto lo hace, por un lado, mejorando y clarificando argumentos que ya habían aparecido en otras partes; por el otro lado, realizando varias movidas originales, algunas de las cuales son extremadamente penetrantes y, por último, por la magnitud misma de la tarea. Creo que la filosofía de la mente que se expone en este

* Publicado en *The Philosophical Review* 79 (1970).

trabajo tiene ciertos defectos, que son inherentes al intento de describir al materialismo como una doctrina no perturbadora. Antes de embarcarme en la crítica, quisiera presentar esquemáticamente esta posición.

Para Armstrong, la tarea por realizar se establece del siguiente modo:

> El objeto que llamamos "cerebro" se denomina así en virtud de algunas características físicas: se trata de cierta clase de objeto físico que se encuentra dentro del cráneo de la gente. Sin embargo, si decimos que este objeto es, además, la mente, entonces, ya que la palabra "mente" no quiere decir lo mismo que la palabra "cerebro", parece que el cerebro sólo puede ser la mente en virtud de alguna característica *adicional* del cerebro. (p.78)

Armstrong concluye que esta característica adicional, o serie de características, debe ser también física, si es que el materialismo es verdadero. La solución consiste en hacer un análisis causal de los conceptos mentales, según el cual se seleccionan "estados de la persona, aptos para causar cierto tipo de conducta". (A fin de evitar circularidad, esta debe ser "conducta puramente física", más bien que "propio de la conducta".) En segundo lugar, algunos estados mentales son también estados de la persona, aptos para ser causados por cierto tipo de estímulos. La palabra "apto" cubre aquí muchas relaciones.

Un rasgo esencial de este análisis es que no nos dice nada acerca de la naturaleza intrínseca de los estados mentales. Estos caen bajo el concepto de la mente, debido a su papel de intermediario en la cadena causal entre estímulos y respuestas, pero más allá de ahí los conceptos mentales no nos dicen nada sobre sí mismos: son tópicamente neutrales [*topic-neutral*]. Por lo tanto, si se puede llevar a cabo el análisis causal puramente en términos de estímulos físicos y conducta física, quedará despejado el camino para la identificación de los estados, procesos y eventos especificados de este modo con los estados, procesos y eventos del sistema nervioso central. Y no se presupondrá ningún atributo no físico adicional al aplicar los conceptos mentales. La identificación resultante será en muchos aspectos similar a la identificación del gen con la molécula de ADN.

La verdad del materialismo puede ser sólo establecida por los neurofisiólogos, a los que remite Armstrong frecuentemen-

te a lo largo del libro. Pero la posibilidad del materialismo exige que un análisis reduccionista de los conceptos mentales sea correcto. Una reducción completa de términos mentales, como lo hace el conductismo, es inaceptable, porque no mantiene la posibilidad lógica de las mentes incorpóreas que, según Armstrong, una teoría correcta debería permitir. Su análisis causal admite esa posibilidad manteniendo la conexión entre mente y conducta, porque el análisis no implica que las causas de la conducta sean físicas; por lo tanto, deja abierta la posibilidad de que sean fenómenos de algún tipo que pueden existir separados del cuerpo. Para ser precisos (aunque Armstrong no lo explica de este modo), el análisis tópicamente neutral permite que sea lógicamente posible que lo que cae bajo un concepto mental pueda ser algo cuya existencia incorpórea sea posible. Pero, naturalmente, es también lógicamente posible, según Armstrong, que los estados mentales resulten ser fenómenos cuya existencia incorpórea no sea posible.

El primer tercio del libro plantea dificultades para diversas teorías de la mente alternativas, y esboza el caso del materialismo del estado central [*central-state*]. Incluye un largo ataque a la incorregibilidad, que Armstrong considera fatal para su posición. Casi todo el resto del libro está dedicado a la Segunda Parte, "El concepto de mente", que es considerado el núcleo intelectual del libro. Aquí, Armstrong hace análisis causales de una enorme cantidad de conceptos mentales. Muchos de ellos son reelaboraciones causales de análisis conductistas o disposicionales. Así, Armstrong dice:

> Podemos ser muy optimistas de antemano respecto del éxito de nuestro programa, porque habremos heredado todo el sorprendente progreso realizado por el conductismo analítico, al revelar la naturaleza de los conceptos mentales sin tener que aceptar la doctrina que representó la caída del conductismo, a saber, el rechazo de los estados mentales internos. (p.118)

Gran parte de la teoría resultante es impresionante, aunque tiendo a pensar que es un error que intente ser tan abarcativa. (En la sección sobre emociones, por ejemplo, la vergüenza se despacha en seis renglones; el desagrado, en cinco, y la desesperación, en tres.)

Una noción a la que Armstrong le dedica gran atención es la de propósito. Es fundamental, no sólo para su explicación de

la voluntad, sino también para su explicación de la percepción. El modelo de conducta dotada de propósito es el control de retroalimentación de un cohete autodirigido, con la importante cualidad de que la causa de retroalimentación sea información suministrada por la percepción u otra forma de aprehenderlo. A su vez, la percepción es analizada en términos de la capacidad para la conducta discriminatoria. Armstrong sostiene que las percepciones no son la *base* de nuestros juicios perceptivos, sino simplemente las adquisiciones [*acquirings*] de ellos. Sus análisis causales dependen de que provean una precondición necesaria para la conducta discriminatoria adecuada si el organismo quiere hacer algo donde eso sea necesario. Pero entonces nos encontramos con una amenaza de circularidad, ya que, por un lado, la conducta discriminatoria a la que se refiere el análisis de la percepción debe estar dotada de propósito y, por el otro lado, la conducta dotada de propósito se distingue de un ajuste meramente mecánico (como la alteración de la temperatura corporal), debido a su dependencia de la percepción en tanto causa de retroalimentación. La sugerencia de Armstrong para resolver este problema es que los dos conceptos, percepción y propósito, no pueden ser aplicados en forma independiente: ambos pueden ser aplicados conjuntamente sólo cuando el mecanismo de retroalimentación es enormemente complejo, y la cantidad, sutileza y alcance temporal de la información a la que responde la conducta deliberada genuina, así como la complejidad de las modificaciones conductuales resultantes, permiten establecer una distinción entre este y otros mecanismos de retroalimentación más simples, que habitualmente no caracterizamos de la misma manera.

Armstrong describe el capítulo 11, en el que desarrolla esta última posición, como el capítulo central del libro. Pero creo que su contribución filosófica más importante y original aparece en el siguiente, referido a las cualidades secundarias. El describe su posición como un reduccionismo realista, y es mucho más convincente que el reduccionismo conductista desarrollado por J. J. C. Smart para acomodar los mismos conceptos recalcitrantes a una visión materialista. Armstrong simplemente utiliza para el análisis de las cualidades secundarias el mismo instrumento que utilizó para el análisis de los estados mentales, analizándolos como estados que poseen ciertas propiedades causales externas, pero cuya naturaleza intrínse-

ca no está especificada en esta identificación. La propiedad de la rojez, por ejemplo, es común a varios objetos y sustancias familiares, que los observadores normales pueden detectar por medio de los ojos. Además, es uno de los determinados que caen bajo un único determinable, en un orden familiar con otros colores. Pero este análisis del concepto no tiene implicaciones sobre cuál es la propiedad o ese determinable, así como el análisis causal de un estado mental no implica nada sobre la naturaleza intrínseca de ese estado. Y, justamente, así como el estado mental puede ser contingentemente idéntico a un estado físico del sistema nervioso central, así el color u otra cualidad secundaria puede ser contingentemente identificado con alguna propiedad física de objetos o sus superficies si descubrimos qué es responsable en nuestra capacidad para hacer discriminaciones y clasificaciones relevantes, con sólo mirar.[1] Más adelante, Armstrong aplica el mismo método de análisis a las sensaciones corporales. El dolor, por ejemplo, es descrito como la percepción de una alteración corporal que evoca típicamente un deseo de que la percepción cese, pero sin que esto implique algo más sobre la naturaleza intrínseca de la alteración, de modo que puede ser "contingentemente" identificado con la estimulación de receptores de dolor.

A la luz de su análisis de las cualidades secundarias, no entiendo las observaciones de Armstrong sobre la posibilidad de los espectros invertidos, que es algo que él quiere defender. Dice que la mitad de la población puede tener un estado interno cuando mira objetos rojos, y otro estado interno cuando mira objetos verdes, en tanto que con la otra mitad ocurre la inversa. Y sostiene (p. 258) que, aun si sus conductas no difieren, al menos uno de los grupos tendría creencias erróneas sobre el mundo. Pero ¿qué creencias erróneas serían? En el análisis de Armstrong, el contenido de los juicios de percepción ordinarios no se vería afectado si ambos grupos fueran afectados de estas diferentes maneras por la propiedad común de los objetos rojos. El juicio "esto es rojo" recibiría el mismo análisis para ambos grupos. Si esto es cierto, ¿cómo podría un grupo estar en lo correcto, y el otro, estar equivocado? Más aún, si Armstrong está en lo cierto al sostener que las percepciones son meramente la adquisición de juicios perceptivos, entonces ambos grupos no tendrán diferentes percepciones, aunque la misma percepción sea un estado físico diferente en cada grupo.

No es posible aquí discutir la mayoría de los análisis que Armstrong ofrece. Los estados mentales se explican en términos de muchas relaciones diferentes con la conducta, incluyendo causa directa, condición necesaria y meras semejanzas con otros estados mentales que se encuentran en relación causal con la conducta. Armstrong analiza la introspección como la percepción interna de los estados mentales propios. Así como la percepción externa, las manifestaciones de conducta de la instrospección involucran la capacidad de discriminar, controlar y tomar en cuenta las circunstancias de las que es percepción, siendo éstas los estados de la persona aptos para la producción de una determinada conducta.

El libro contiene un análisis del conocimiento basado en (a) una teoría causal de la inferencia y (b) una definición de conocimiento no inferencial que utiliza la noción de una creencia que es *empíricamente suficiente* para su propia verdad. Excepto por agregar la condición *"en esta situación particular"*, Armstrong abandona la noción de suficiencia empírica no explicada, por lo que la encuentro opaca. Por ejemplo, ¿implica algo sobre la relación causal (en cualquier dirección) entre la creencia y la circunstancia que la hace verdadera? ¿Cómo se define "esta situación particular"? Ya que la situación en sí incluye la verdad de la creencia, es necesaria una interpretación. A pesar de la extensión del libro, su esfuerzo por abarcar tanto acarrea, a veces, un tratamiento deficiente de temas importantes.

II

Ahora quiero plantear algunas cuestiones generales sobre el tipo de materialismo que Armstrong defiende como posible, y sobre el análisis causal de los conceptos mentales. Mi primer problema es el siguiente: ¿por qué una teoría mentalista del funcionamiento de los seres humanos debería corresponderse de cerca con la imagen mentalista, a fin de permitir la identificación de ítems de ambas teorías? Aun si alguna forma del materialismo es verdadera, no será automáticamente expresable en el marco de la psicología de sentido común. Los datos de que disponemos actualmente sobre el sistema nervioso central no parecen darnos esperanza en ese sentido, y algunos, por el

contrario, lo desalientan. Por ejemplo, la bifurcación fundamental izquierda-derecha de la función cerebral, a la que no corresponde nada en la psicología de la percepción de sentido común.

Hay un aspecto importante en el que fracasa la analogía de Armstrong entre el materialismo y la ecuación gen-molécula de ADN. La genética premolecular era una teoría científica exacta: se introdujo el concepto de gen al servicio de la explicación científica, y el trabajo posterior en la biología molecular siguió ese mismo objetivo. Pero la psicología de sentido común, organizada con los conceptos ordinarios de creencia, deseo, sensación, percepción, emoción, etc., no es una teoría científica. Los estados mentales de los que Armstrong hace un análisis causal se eligen por un sistema que evolucionó naturalmente, cuya forma puede depender de manera significativa de funciones extracientíficas. En nuestras relaciones y comunicaciones mutuas necesitamos un vocabulario especializado y, aunque nos sirva bastante bien para la vida cotidiana, su estrechez e inadecuación en tanto teoría psicológica se vuelven evidentes cuando intentamos aplicarlo a la formulación de descripciones generales de la conducta humana, o para la explicación de condiciones mentales anormales.

No deberíamos esperar que la teoría causal cruda e incompleta contenida en la psicología de sentido común sobreviva intacta los próximos cien años de investigación sobre el sistema nervioso central. Sería sorprendente que conceptos como creencia y deseo encontraran algo que les correspondiera en una teoría neurofisiológica, considerando cuán limitado es su poder explicativo y predictivo, incluso para la conducta más general. La conducta física de que, según el análisis de Armstrong, una intención dada es apta para causar puede ser el producto de causas cuya complejidad no podemos poner ni siquiera en una correspondencia aproximada con los elementos simples de la explicación psicológica actual.

Si esto es así, entonces una teoría fisicalista del funcionamiento humano no tomará la forma de identificaciones entre estados psicológicos a la antigua y estados físicos del sistema nervioso central descritos en forma microscópica. En cambio, adoptará los conceptos de alguna teoría más adelantada del funcionamiento humano superior. Más aún, no podemos suponer que estos conceptos se extraigan de alguna de las teorías

actuales de la descripción del cerebro. La neurofisiología puede descubrir fenómenos con los cuales no estamos familiarizados y que no se reducen a las múltiples ocurrencias de fenómenos con los que sí estamos familiarizados. (Armstrong reconoce esta posibilidad, e incluso aquella de que el materialismo que resulte correcto pueda no ser un materialismo fisicoquímico.) Por lo tanto, dudo de que los términos para expresar ambos aspectos de una identidad fisicalista sean accesibles hoy en día, y el desarrollo de la psicología fisiológica podría dejarnos tan atados a una teoría común que cualquier identidad que tratáramos de formular sería tautológica. Nada de esto es una objeción al materialismo, pero sugiere que la formulación de esa doctrina necesita mucho para progresar más allá de los términos de la teoría de la identidad tradicional.

Ahora quisiera ver cuál es el alcance exacto de los análisis causales de Armstrong, y si es que tienen éxito en alcanzar sus objetivos. En las páginas 84 y 85, dice que no intenta hacer *traducciones* de estados mentales, sino sólo *hacer justicia* a la naturaleza de los estados mentales por medio de conceptos puramente físicos o neutrales. Pero en la página 90 se refiere a esta empresa como un análisis lógico de los conceptos mentales y como una tesis conceptual. Asumo que él considera que sus análisis son parciales, es decir, indicaciones aproximadas de las clases de fenómenos físicos a las que nuestros conceptos mentales están atados. De hecho, Amstrong se apoya, en gran parte, en nuestro conocimiento no verbal cotidiano de las características conductuales de los seres conscientes, y no hace un análisis físico muy detallado. Pero no hay necesidad, para sus propósitos, de dar un análisis en términos físicos que sea *independiente* del vocabulario psicológico. Si los conceptos mentales se aplican a condiciones causalmente relacionadas con cierta conducta, es improbable que tengamos el vocabulario para una descripción independiente de precisamente esa conducta, sin referencia a la clase de estado mental que manifiesta. No se puede suponer, en general, que un rasgo físico al que respondemos en forma uniforme, y cuyo nombre podemos aprender, sea analizable en términos de rasgos físicos componentes, para los cuales ya tenemos (o podríamos tener) un vocabulario. La mayoría de nuestros conceptos no artificiales pueden ser primitivos, en este sentido.

Pero, si esto es así, entonces no tenemos que buscar un *análisis* de los conceptos mentales en otros términos. La teoría

causal de Armstrong puede enunciar verdades sobre nuestros estados mentales, aun si no nos proporciona análisis adecuados de los conceptos mentales. Un materialismo del tipo que quiere Armstrong requeriría que los estados mentales *participaran de hecho* en la causación de la conducta. Pero, respecto del *análisis* de la mente, hay que establecer sólo un requisito negativo: que los conceptos mentales no impliquen la inclusión de ninguna sustancia o atributo no físico.

Sin embargo, incluso la defensa de Armstrong de esta tesis negativa acarrea serias dificultades, llevándolo a la poco común afirmación de que la naturaleza intrínseca de nuestros estados mentales no desempeña ningún papel en nuestros conceptos mentales, y que, por lo tanto, no podremos conocerla hasta que descubramos cuáles son las causas de nuestra conducta. Un rasgo sorprendente del libro de Armstrong es que él no piensa que esta sea una afirmación que encontremos difícil de aceptar. Nunca toma en serio la objeción natural de que debemos conocer la naturaleza intrínseca de nuestros *propios* estados mentales, dado que los experimentamos directamente.[2]

Consideremos, por ejemplo, su tratamiento de la posibilidad de concebir la existencia incorpórea. Armstrong cree que el elemento tópicamente neutral de su análisis causal resuelve este punto, ya que permite la posibilidad de que las causas de nuestra conducta puedan ser no físicas. Pero nótese que esto no es lo mismo que admitir la posibilidad de la existencia incorpórea: es sólo admitir la posibilidad de esa posibilidad. No se trata de una distinción trivial. Armstrong se contenta con sostener que no sabemos que la existencia incorpórea de la mente es inconcebible. Pero algunos filósofos sostendrían que sabemos que la existencia incorpórea de la mente *es* concebible. Esta última es la tesis más fuerte para la prueba de Descartes de que la mente y el cuerpo son distintos. El no consideraba meramente que la mente podía ser algo distinto del cuerpo y por lo tanto capaz de tener una existencia independiente. Más bien, sostenía que podemos concebir sin dificultad su existencia independiente, y que, por lo tanto, debería ser algo distinto del cuerpo.

Este es un argumento poderoso. Si queremos oponernos a él, debemos sostener, como Armstrong, que Descartes, al estar sujeto a sus propios estados mentales, no es consciente de su naturaleza intrínseca, por lo que no tiene ninguna base para sos-

tener que puede concebir que, por ejemplo, *esta clase de sensaciones* ocurren aunque su cuerpo no exista. Ciertamente, a Descartes no le gustaría que le dijéramos que él no puede estar seguro de que la ocurrencia incorpórea de esas sensaciones es *in*concebible, ya que él está seguro de que *es* concebible, ya que las ha concebido.

El argumento de Descartes también tiene la siguiente forma inversa que, a mi entender, él nunca utilizó: la existencia del cuerpo sin la mente es tan concebible como la existencia de la mente sin el cuerpo. Es decir, puedo concebir mi cuerpo haciendo justamente lo que está haciendo ahora, interior y exteriormente, con causación física completa en su conducta (incluyendo la conducta típicamente autoconsciente), pero sin incluir ninguno de los estados mentales que estoy experimentando en este momento, ni ningún otro. Si esto fuera realmente concebible, entonces los estados mentales deberían ser distintos de los estados físicos del cuerpo, y el análisis causal de Armstrong de los conceptos mentales no podría ser correcto, ya que según ese análisis la presencia de causas adecuadamente complejas de conducta adecuadamente compleja *implicarían* la existencia de estados mentales conscientes.

Los ejercicios conceptuales de los que dependen estos argumentos son muy convincentes. Si Armstrong está en lo correcto, son ilusorios. Pero, entonces, ¿qué estoy haciendo cuando concibo a mi mente sin mi cuerpo, o viceversa? Por cierto, no estoy meramente imaginando que las causas de mi conducta corporal son causas de una sustancia no física que puede ser concebida separadamente del cuerpo. El concepto de posibilidad aquí no tiene nada que ver con el de analiticidad. Lo que se sostiene aquí no es que bajo cierta descripción (a saber, una descripción mental) no hay contradicción en suponer que mis estados mentales deben existir separados de mi cuerpo, sino, más bien, que en virtud de lo que ellos *son*, es concebible su separación del cuerpo.[3] El problema real es si podemos saber que hemos concebido tal cosa, y si la familiarización con los contenidos de nuestra mente nos coloca en una mejor posición para hacerlo que concebir que el lucero del alba se mantiene mientras que el lucero vespertino se destruye. (Lo cual es inconcebible, ya que son lo mismo. Su posibilidad de ser concebido no puede ser probada mostrando que no es una contradicción, pues sólo mostraría que esas descripciones no *implican*

que es inconcebible.) Para refutar el argumento cartesiano, debemos mostrar que cuando tratamos de concebir nuestras mentes sin los cuerpos, o viceversa, no podemos hacerlo y, en cambio, hacemos otra cosa que pensamos que es eso. Por ejemplo, podemos concebir que otros seres son psicológicamente similares a nosotros, pero con otra constitución psicofísica; lo cual no sería concebir que nuestros propios estados mentales funcionan sin el tipo de cuerpo que tenemos ahora. O bien Descartes está equivocado al pensar que podemos concebir su existencia separada, o bien algún tipo de dualismo es correcto.

Un indicio de que algo anda mal es que Armstrong considera que es una *virtud* de su teoría despachar rápidamente el problema de otras mentes. Dice:

> Supongamos que un cuerpo humano exhibe la clase de conducta correcta. Dado nuestro análisis de la naturaleza de los estados mentales, necesitamos sólo tres premisas para inferir la existencia de la mente, de la cual esta conducta es una expresión. (i) La conducta tiene *alguna* causa; (ii) la causa se encuentra en la persona que tiene la conducta; (iii) la causa es "adecuada", es decir, tiene la complejidad que corresponde a la complejidad de la conducta en cuestión. Dadas estas suposiciones bastante modestas, es necesaria la existencia de otra mente. (pp. 124-125)

Pero esta "solución" al problema deja en el misterio la razón por la que, más de una vez, ha producido molestias. El problema surge precisamente en el punto en que las tres premisas de Armstrong son satisfechas, y descubrimos que la existencia de la mente puede igual ser puesta en duda. Una teoría de la mente que desecha la duda sin dar cuenta de su origen deja inconclusa buena cantidad de trabajo filosófico serio.

Sospecho que estas no son dificultades del materialismo *per se*, sino, más bien, que surgen de rasgos de la aplicación de conceptos mentales en primera persona, que no son dados a conocer en el análisis causal directo de Armstrong, lo cual puede referirse a la utilidad de la teoría de Armstrong en tanto una formulación del materialismo. Pero también está la cuestión de si se prueba que nuestros conceptos mentales cotidianos, como sea que los analicemos, tienen alguna conexión exacta con una teoría física de la operación de los organismos que describimos como humanos. Si no, es posible que debamos reexaminar la suposición de que somos personas.

Notas

1. En la pág. 275, Armstrong sugiere que el análisis del concepto de rojez nos da sólo verdades *contingentes* sobre la rojez. Aunque no lo diga exactamente de este modo, se puede agregar que lo que es necesariamente verdadero sobre la rojez, su naturaleza intrínseca, puede descubrirse sólo con investigación empírica. Esto sugiere la posibilidad de un ataque más general a la asociación tradicional de lo necesario con lo analítico y el conocimiento a priori. De hecho, esta investigación fue realizada por Saul Kripke, pero sus resultados no han sido publicados. [Luego fueron publicados en *Naming and Necessity* (Harvard University Press, 1980).] Estos estudios sugieren, sin embargo, que no deberíamos considerar al materialismo del estado central simplemente como una tesis contingente, sino que deberíamos decir, en cambio, que, si la versión materialista de Armstrong es verdadera, es necesariamente verdadera de los estados mentales de los que ellos son estados físicos, pero solo contingentemente verdaderos de los que ellos son los estados mentales, ya que esto último depende de sus efectos más que de su naturaleza intrínseca. Lo que es analíticamente v dadero de algo, bajo cierta descripción, puede no ser necesariamente verdadero de ello, ya que la descripción puede adecuarse a la cosa sólo contingentemente. Por el otro lado, un enunciado no analítico sobre ello puede ser necesariamente verdadero si, por ejemplo, dice qué es la cosa.

2. Aunque su ataque a la incorregibilidad presupone un rechazo a la idea de un contacto directo con nuestros estados mentales propios: él sostiene que sólo los conocemos por descripción.

3. Nuevamente, debo mencionar el trabajo de Kripke sobre posibilidad y necesidad, que me condujo a ver que hay más que una confusión entre sinonimia e identidad en el argumento cartesiano, que se basa en el reconocimiento de que si un evento mental *es* un evento físico, entonces no es posible que aquel ocurra independientemente de este.

7

Dennett: Contenido y conciencia*

Este ensayo fue una reseña del primer libro de Daniel C. Dennett, Contenido y conciencia *(Routledge and Kegan Paul / Humanities Press, 1969). Así como la mayoría de los escritos de Dennett, el libro me pareció sumamente agradable y lleno de información científica fascinante, pero nada convincente con respecto a su tesis principal.*

Aunque no resuelve el problema mente-cuerpo, se trata de un libro valioso e interesante. Dennett comienza con la afirmación verdadera de que nadie encontraría aceptable la teoría de la identidad, si no fuera por el poco atractivo de las alternativas. Tal como él lo expresa, la razón de la identidad es evitar el bulto óntico. Si realmente hay fenómenos mentales, deben ser idénticos o no a los fenómenos físicos. Si debemos evitar el dualismo, y si el conductismo no es un camino viable para evitarlo, parece que caemos en los brazos de una teoría de la identidad neural.

El intento de Dennett de mostrar que tal teoría no puede ser correcta padece de su suposición de que las identidades deberían ser generales, más bien que entre sucesos mentales particulares y sucesos neurales particulares, una suposición que no todos los defensores de la teoría aceptan. Sin embargo, Dennett tiene razón con respecto a la motivación negativa, que es su apoyo más poderoso, y que ha llevado a los filósofos a gastar tanta tinta y energía para hacerla funcionar. Por ello, su promesa de encontrar una nueva alternativa merece un interés considerable.

* Reimpreso con autorización del *Journal of Philosophy* 69 (1972).

Dennett sugiere que los fenómenos mentales no pueden ser ni idénticos ni no idénticos a los fenómenos neuronales, ya que la mayoría de los términos mentales son no referenciales. Se trata de una propuesta radical, ya que los términos como "dolor", "pensamiento", "imagen mental" y "creencia" no refieren, ya que, en cierto sentido, no existen tales cosas.

El ejemplo estándar de un término no referencial es el vocablo inglés "sake". No hay "sake", pero hay verdades susceptibles de ser enunciadas en términos de "sake". Los términos mentales no pueden ser encasillados en expresiones únicas, como ocurre con "sake", pero Dennett analiza el término "voz", como un ejemplo intermedio. Las voces no existen realmente, dice, porque podemos describir las circunstancias en virtud de las cuales cualquier enunciado que contiene la palabra "voz" es verdadero, sin referirnos a una voz (con este o aquel nombre). En cambio, nos referimos a sonidos, cuerdas vocales, variedad de frecuencia, etc. No se trata aquí de dirigir la atención a una diferencia de categoría, sino que se trata de una reducción sin identificación. "Nuestras discusiones sobre categorías o errores categoriales no podrán mantenernos a salvo de las trampas del dualismo, a no ser que estemos preparados para garantizarle prioridad ontológica a una categoría a expensas de la otra." (p. 13)

Dennett argumenta que un método de análisis similar podría ser útil para el vocabulario mental. Si pegamos los términos mentales en sus contextos y permanecemos ontológicamente neutrales con respecto a sus referencias, podemos buscar una "explicación científica de las diferencias y semejanzas de lo que es el caso, en virtud del cual son verdaderas o falsas diferentes oraciones del lenguaje mental". (p. 18) "Nuestra tarea implica, por lo menos, lo siguiente: enmarcar dentro del lenguaje científico los criterios —las condiciones necesarias y suficientes— de la verdad de las oraciones del lenguaje mental." (p. 19) Presumiblemente, no se supone que estas condiciones de verdad den el *significado* de las oraciones del lenguaje mental, y no me queda claro exactamente qué tipo de condiciones se supone que son, y cómo podemos saber si han sido descritas correctamente. Ya que este nuevo método de análisis es el corazón del libro, habría sido adecuado analizar más profundamente su carácter lógico.

Luego de formular la tarea por realizar, Dennett declara que el principal obstáculo aparente para llevarla a cabo es la

intencionalidad de la esencialidad mental. Básicamente, considera el problema mente-cuerpo como un caso especial de la pregunta de cómo un mecanismo físico puede ser un sistema Intencional, lo cual es, a mi entender, un error. Podemos responder la pregunta, dice Dennett, si encontramos una forma de atribuirles significación a los estados internos del mecanismo. Podemos hacerlo con ordenadores, lo cual sería parasitario respecto de la Intencionalidad de los usos que los seres humanos les dan a los ordenadores: los estados internos de un ordenador no tienen ninguna significación por sí mismos.

Sin embargo, la significación de los estados cerebrales no se explicaría del mismo modo, sin retroceder un paso en el problema. La explicación no puede poner furtivamente un hombrecito en el cerebro, que les dé la significación a los estados cuando los interpreta. Dennett sugiere que el tipo deseado de explicación no antropomórfica puede ser formulado con referencia a la evolución y selección natural, que apoyan las atribuciones de función, y podrían apoyar atribuciones más complejas de significación si se aplicaran a los estados de un organismo evolucionado que responda de maneras complejas a la información interna y externa.

Lamentablemente, al demostrar cómo podría realizarse esta tarea, Dennett revela una fuerte veta conductista. Como ejemplo del tipo de análisis en el que piensa, trata el caso del dolor. En el nivel que él llama *personal*, el dolor explica una conducta de prevención y angustia: aquí es donde se detiene la explicación. Dar una explicación fisiológica de los mismos sucesos significa abandonar el nivel personal, pasando al nivel *subpersonal*. Sin embargo, es posible, en el nivel subpersonal, explicar la significación de la conducta ordinaria de dolor. (Dennett se imagina el proceso por el cual tales respuestas compulsivas al daño o a la amenaza de daño podrían haber evolucionado.) Pero si una explicación de *esto* es todo lo requerido para describir en lenguaje científico esos hechos en virtud de los cuales es verdad que sentimos dolor, entonces la posición es desvergonzadamente conductista. En realidad, Dennett pregunta:

> ¿Podría tener algún sentido la suposición de que una persona podría golpearse el pulgar con un martillo, y sentirse repentina e imperiosamente forzada a tirar el martillo, chuparse el pulgar, dar saltos, gritar, gemir, llorar, etc. y, sin embargo, no experimentar

> dolor? Es decir en este caso no estaría actuando como en un escenario: estaría forzada a actuar de ese modo. Sería incapaz de responder al aplauso con una reverencia sonriente. Suponer alguna cualidad o fenómeno horrible (pero imposible de ser descrito en otros términos) para acompañar dicha conducta forzada sería completamente gratuito. (pp. 94-95)

Se trata de una palabra curiosa. Podría ser gratuito, pero ciertamente hay una *diferencia* entre este tipo de conducta con dolor y esta misma conducta sin dolor. Un organismo que pone en marcha la conducta compulsiva preventiva del dolor, pero sin ese sentimiento interno, sería perfectamente imaginable, y la explicación subpersonal que provee Dennett podría aplicársele perfectamente, así como se nos puede aplicar a nosotros. Por lo tanto, aunque él haya descrito con lenguaje científico las bases de ciertos rasgos Intencionales o funcionales del dolor, no ha aprehendido lo que es el dolor en mayor medida que cuando alguien que describe las condiciones bajo las que algo es un nombre, nos define lo que es la tinta. Dennett dice, naturalmente, que la palabra "dolor" no tiene referente concreto. Pero no puede sostenerlo hasta tanto no haya abarcado exhaustivamente las condiciones de verdad de los enunciados acerca del dolor, sobre la base del modelo anterior de disección de voces.

Luego, sigue un ataque semejante al fenómeno de la conciencia; y, nuevamente, resulta que lo que se explica en el nivel subpersonal no es la conciencia, sino sólo ciertos patrones de conducta que son característicos de la conciencia, pero compatibles con su ausencia. Dennett sugiere analizar la conciencia, una noción compleja, en términos de dos tipos de toma de percatación:

> (1) A se da cuenta$_1$ de que *p* en un tiempo *t,* si y sólo si, *p* es el contenido del estado de entrada del "centro del habla" en el tiempo *t.*
>
> (2) A se da cuenta$_2$ de que *p* en un tiempo *t,* si y sólo si, *p* es el contenido de un suceso interno en A en el tiempo *t,* que es eficaz para dirigir la conducta habitual. (pp. 118-119)

Dennett desarrolla aun más estas definiciones en términos de almacenamiento y recuperación de información. Reconoce que la toma de conciencia$_1$ puede aplicarse a ciertas máquinas.

Un análisis de la conciencia de este tipo puede conducirnos a una solución del problema mente-cuerpo respecto de máquinas, pero este es un tema diferente del problema tradicional mente-cuerpo. Dennett sostiene que los diversos casos que se agrupan bajo la palabra ordinaria "conciencia" consisten en la capacidad de darse cuenta$_1$ o darse cuenta$_2$ en diferentes grados, y no hay lugar "para un concepto de toma de conciencia$_3$ que se aplique sólo a personas, dejando afuera todas las máquinas imaginables" (p. 121). Pero Dennett no hace esta afirmación: ni siquiera la defiende. Porque supone que la intencionalidad de los estados mentales es el principal obstáculo para escapar del dualismo. Las alternativas de Dennett al dualismo dejan intacto el núcleo del problema.

Más adelante, Dennett analiza percepción, imaginación, pensamiento, acción e intención. Su motivación es el objetivo de dar cuenta de estos fenómenos de modo que no reintroduzcan simplemente al analizando dentro del análisis asumiendo, de un modo u otro, algún hombrecito o comité de hombrecitos en el cerebro. Dennett observa que:

> Si una imagen debe funcionar como un elemento en la *percepción*, debe funcionar como la materia bruta, no como el producto final, ya que, si suponemos que el producto del proceso perceptivo es una imagen, deberemos designar algún perceptor análogo que esté sentado frente a la imagen, y a otro que esté sentado frente a la imagen que es su producto final de la percepción del perceptor-análogo, y así sucesivamente *ad infinitum*. (p. 134)

Mucho de lo que Dennett dice sobre los conceptos mentales es imaginativo y convincente. Pero su tarea fundamental es reducir nuestro compromiso ontológico de lo mental, explicándolo en el nivel subpersonal, "donde la gente, los pensamientos, experiencias e informes introspectivos simplemente no forman parte del tema en cuestión" (p. 113), como puede haber verdades no referenciales acerca "de" estas cosas en el nivel personal. Esta tarea permanece incompleta, y sus condiciones de éxito, dudosas. Pueden decirse muchas cosas que sean verdades sobre la mente y su relación con el cuerpo, dejando el problema mente-cuerpo fundamentalmente inalterado.

8

Dennett: La disolución de la conciencia*

Este ensayo fue una reseña del libro de Daniel C. Dennett, Consciousness Explained *(Little, Brown, 1991). Es de notar cuán poco ha cambiado la opinión de Dennett, y, naturalmente, la mía tampoco. La opinión de Dennett es, básicamente, una combinación de Gilbert Ryle con el* Scientific American: *un eterno optimismo sobre la posibilidad de aclarar el problema mente-cuerpo mediante algunas observaciones razonables sobre los criterios de conducta de las atribuciones mentales y, además, muchos hechos empíricos. Sigo pensando que él pierde de vista, desde el comienzo, el problema real. Creo que nuestro desacuerdo sólo podrá terminarse en nuestra tumba, si acaso alguna vez se termina.*

La más profunda división entre los que escriben sobre la mente y el cerebro se encuentra entre reduccionistas y antirreduccionistas: entre quienes piensan que, algún día, los fenómenos mentales podrán ser explicados con los recursos de la ciencia física, y quienes piensan que los fenómenos mentales son tan radicalmente diferentes de todo lo demás que se necesita una extensión igualmente radical de las formas de explicación científica para explicar cómo ellos surgen de la operación física del cerebro.

En esta discusión la experiencia consciente ocupa un lugar central, porque uno de los desafíos más poderosos para el

* Reimpreso con autorización de *The Wall Street Journal*. Copyright © 1991 por Dow Jones & Company. Todos los derechos reservados. Publicado por primera vez el 7 de noviembre de 1991.

reduccionista es explicarle al antirreduccionista cómo es concebible que el gusto subjetivo del chocolate, por ejemplo, no está meramente *causado* por ciertos estímulos neuronales en el cerebro que, a su vez, están causados por estímulos en las papilas gustativas, sino que consisten en *nada más* que en estos u otros eventos susceptibles de ser descritos físicamente. La ciencia física fue desarrollada, después de todo, para tratar con lo que está fuera de la mente, independientemente de cómo lo percibamos.

Daniel C. Dennett es un defensor prominente e incansable de la posición reduccionista, la posición de la mayoría de los que trabajan en este tema. *Consciousness Explained* es un ataque a gran escala al problema de la conciencia. Es, para usar sus propias palabras en otro de sus trabajos recientes, un fracaso instructivo: instructivo, tanto en sí mismo como en sus fracasos por encontrar una explicación convincentemente completa de la conciencia a partir de los materiales limitados que él mismo se permite usar.

El libro es un ejercicio de teorización utópica sobre la base de la información disponible, es decir, una actividad admirable en un tema tan poco desarrollado como este. Sus explicaciones no se ofrecen como hipótesis comprobables, sino como modos de pensar acerca de la relación mente-cerebro que colocarían a la conciencia confortablemente en un mundo, cuyos componentes básicos fueran completamente susceptibles de ser descritos por la física, mostrando que "los diversos fenómenos que componen lo que llamamos conciencia... son todos ellos efectos físicos de las actividades del cerebro". La analogía de la relación software-hardware en los ordenadores desempeña un papel fundamental en esta historia, y Dennett hace un gran esfuerzo, dada su falta de simpatía, para explicar los argumentos antirreduccionistas principales, y para destrozarlos con la actitud de conjurar nubes de ilusión y superstición.

Dennett es un escritor inteligente y dotado, y el libro está lleno de información fascinante sobre los humanos, animales y máquinas. El resultado es sumamente fácil de digerir y un útil paseo por el tema, pero al final se llega a la conclusión de que lo que Dennett quiere hacer no puede hacerse.

Su proyecto está marcado por la siguiente suposición metodológica, que él simplemente enuncia, pero que debería haber sido un foco central de su argumento: "dado que no po-

demos ver directamente el interior de la mente de la gente, sino que tenemos que creer en su palabra, cualquiera de los hechos que tenemos sobre sucesos mentales no se cuentan entre los datos de la ciencia, ya que nunca pueden ser correctamente verificados con métodos objetivos". Y sigue: "El desafío es construir una teoría de eventos mentales, usando los datos que el método científico permite. Una teoría de este tipo deberá construirse desde el punto de vista de la tercera persona, ya que *toda* ciencia se construye desde esa perspectiva".

Con esto, enuncia su estrategia, que él denomina "heterofenomenología". De hecho, el procedimiento descansa implícitamente en nuestro conocimiento de primera persona de la conciencia, aunque pretende prescindir de él. Lo que hace es construir una teoría de tercera persona que, aunque utiliza términos psicológicos familiares del tipo "como si", teoría que podría aplicar a los robots, se basa exclusivamente en las manifestaciones conductuales y fisiológicas por lo general asociadas a la conciencia, sin suponer que los seres de los que habla (en general, humanos) sean conscientes.

Una teoría de este tipo podría ser también aplicable a un autómata suficientemente complejo y animado, pero subjetivamente inconsciente, si es que tal cosa puede existir. Sin embargo, una vez explicada su teoría de la tercera persona, Dennett plantea la pregunta: ¿Podría un sistema físico que satisfaga esta descripción realmente no ser consciente? ¿No faltaría algo para que haya conciencia? Y su respuesta es: No. Eso es todo lo que la conciencia es.

La idea de que los sucesos mentales no se cuentan entre los datos de la ciencia era la premisa que guiaba al conductismo. La posición de Dennett es un derivado sofisticado de esta idea: ambas dependen de la misma confusión respecto de la objetividad. La objetividad en cualquier área de pensamiento requiere algún método de confirmación o no de las observaciones o juicios de un individuo con referencia a los de los demás. Pero la forma particular en que hacemos esto para los datos físicos está determinada por la naturaleza de esos datos y no debería identificarse con la objetividad en general.

Aunque la idea de datos objetivos no sea adecuada para los datos mentales, no quiere decir que carezca de sentido: sólo quiere decir que no necesitamos estándares objetivos que combinen los puntos de vista de la primera y tercera persona, ya

que de hecho están combinados en los conceptos ordinarios para la atribución de estados conscientes que todos utilizamos sin dificultad, para corregir nuestras descripciones experienciales, y las de los demás. Una teoría de la conciencia que no incluya sucesos mentales entre sus datos es como un libro sobre Picasso que no mencione sus pinturas.

Aunque rechacemos la tesis reduccionista amplia de Dennett, su teoría es interesante, dado que gran parte de ella sería verdad respecto de la conciencia aun si no es todo lo que se puede decir sobre la conciencia. Dennett se opone tanto a lo que llama materialismo cartesiano, como al dualismo cartesiano. El cree que no hay ningún "teatro cartesiano" en el cerebro, en el que los contenidos de la conciencia están a la vista de algún ser material. Tal como él dice: "la glándula pineal no sólo no es el aparato de fax hacia el alma, sino que tampoco es el Salón Oval del cerebro, ni tampoco lo son otras porciones de este". De algún modo, la conciencia tiene que abarcar todo el cerebro, pero, lo que es más interesante aún, tiene que estar en el tiempo, es decir, el tiempo físico. Algunas de los mejores análisis del libro tratan resultados experimentales que muestran que la apariencia subjetiva de la secuencia temporal o la simultaneidad temporal entre experiencias conscientes no puede comprenderse en términos de las relaciones correspondientes de sucesos cerebrales en tiempo real. La reflexión sobre el carácter no localizado y no lineal de la conciencia lleva a Dennett a sugerir una imagen compleja del cerebro, como produciendo múltiples borradores de versiones del mundo y sus aspectos, por parte de muchos subsistemas, que están en competencia recíproca. La conciencia es una "máquina virtual", parecida a un ordenador Neumann (que, de hecho, fue diseñado sobre el modelo del flujo de pensamiento consciente), programado a través de la evolución cultural hacia la arquitectura de proceso-paralelo del cerebro. Dennett lo considera "una innovación demasiado reciente para estar fijo dentro de la maquinaria innata".

Esta última es una afirmación extraña: Dennett parece confundir conciencia con *auto*conciencia. La identifica con una capacidad de hacer juicios de segundo orden sobre nuestros propios estados discriminatorios, en donde esos juicios no necesitan ser ellos mismos conscientes, pero son susceptibles de expresarse verbalmente. Una criatura que no sólo puede discriminar cosas rojas de cosas verdes, sino también decir que algo

le parece rojo, tiene percepciones de color conscientes. Ni el juicio de segundo orden, ni la percepción de primera persona deben poseer alguna otra cualidad experiencial intrínseca: no existen tales cualidades, sino sólo juicios, y estos pueden ser explicados en términos de una teoría de tercera persona, sobre el funcionamiento, incluyendo el funcionamiento verbal, del organismo.

Esta definición inadmisible implica que los bebés no pueden tener sensaciones conscientes antes de poder formarse juicios sobre sí mismos. Pero el problema principal es que esta fórmula, entendida exclusivamente en tercera persona en el sentido heterofenomenológico que Dennett le ha dado, no implica para nada a la conciencia del sujeto. Naturalmente, creeríamos que cualquier cosa que funcione física y conductualmente como un ser humano adulto es consciente, pero esta creencia sería una conclusión a partir de la evidencia, antes que únicamente una creencia en la evidencia. Es sólo la concepción rígida de Dennett sobre la objetividad científica la que lo lleva a pensar de otro modo.

9

O'Shaughnessy: La voluntad*

Este ensayo fue originalmente una reseña de The Will: A Dual-Aspect Theory, *de Brian O'Shaughnessy (Cambridge University Press, 1980). Aunque se trata, en su mayor parte, de una descripción del libro, decidí incluirlo en esta selección debido a mi simpatía por su enfoque sobre la relación mente-cuerpo y su defensa del análisis conceptual como un recurso esencial para la comprensión del mundo, en particular el mundo humano en un nivel básico.*

Todo buen filósofo debe encontrar su obsesión, que lo ha de guiar el resto de su vida. La obsesión de Brian O'Shaughnessy se refiere a la más íntima de las relaciones en las que el yo se encuentra respecto del mundo físico o "externo": su relación con esa parte del mundo físico que él puede mover directamente, y del que tiene conciencia inmediata: su cuerpo.

Hay algo de paradójico en describir el cuerpo como parte del mundo externo simplemente porque es un objeto físico, lo que muestra cuán útil debe ser el tema de la acción humana, en tanto una clave para comprender el lugar del yo en el mundo. Es aquí, si es que acaso hay un lugar, donde la división, aparentemente radical, entre mente y materia puede ser sustituida por una conexión interna inteligible. Porque nuestro cuerpo no está fuera de nosotros, aunque es un objeto físico. No nos damos cuenta de él sólo por sus efectos sobre nuestros sentidos (puedo mirar mi mano para ver su posición en el teclado, pero también sé, sin mirar, cuán lejos está mi índice de la letra *t*). Y

* Reimpreso con autorización de *The Times Literary Supplement*, 27 de marzo de 1981.

no movemos nuestro cuerpo empujándolo de un lado al otro (si se me durmió mi brazo izquierdo, puedo moverlo con mi brazo derecho, pero, en ese caso, no lo muevo *con algo*, simplemente lo muevo).

Pero, si mi cuerpo no está fuera de mí, entonces, ¿mi cuerpo es parte de mi mente, o sea, una porción del mundo físico está contenida en mi yo? Por más extraño que parezca, esta es la respuesta de O'Shaughnessy: "Los sucesos que ocurren fuera del cerebro pueden ser partes, incluso partes esenciales, de sucesos psicológicos *experimentados en forma inmediata*". Quizás, incluso cuando pensamos desde dentro hacia fuera, sería inconcebible que algún yo, remotamente similar al nuestro en lo psicológico, no tuviese un cuerpo en este sentido tan fuerte: no es que el yo esté *albergado* en un cuerpo, sino que la acción física directa y el conocimiento directo de su propia postura y orientación física son partes de su vida mental. Y este desgarramiento del yo por el cuerpo en la acción puede afectar los fenómenos psicológicos de sensación, percepción, deseo, creencia, intención, pensamiento, que son más "internos", debido a su conexión cercana con la acción y toma de conciencia, en una única red conceptual.

Esta inversión en la dirección habitual para enfocar el problema mente-cuerpo es el rasgo principal del libro de O'Shaughnessy. El investiga la relación entre mente y cuerpo desde *dentro de la mente misma*, no como aparecería a la observación externa de los efectos de conducta o causas fisiológicas de fenómenos psicológicos. O'Shaughnessy quiere comprender ese aspecto misterioso y esencial de la vida interna de cada uno de nosotros: la condición de *ser* un organismo físico, animal.

A fin de defender este punto de vista, debe oponerse a una imagen rival de la relación íntima entre el yo y el cuerpo: que movemos nuestro cuerpo directamente por el efecto causal de un tipo especial del acto mental interno, una volición, y tenemos conciencia directa de nuestro cuerpo a través de una imagen corporal que es una clase de huella sensorial pasiva. Según esta posición, nuestro cuerpo está mucho más cerca de nuestra mente que cualquier otro objeto físico, pero ambos permanecen esencialmente distintos y su relación recíproca es casi mágica así como sería mágico poder mover los muebles con sólo desear que se movieran. O'Shaughnessy se opone a esto con una imagen no mágica que, sin embargo, no es una reducción

de lo mental a lo físico, y, al hacerlo, propone una teoría general y original sobre la estructura de la mente.

Antes de describir los resultados, deseo decir algo del contexto filosófico. Este trabajo trata el problema mente-cuerpo pero, aunque expresa una deuda con Wittgenstein y con los escritos de Elizabeth Anscombe, Stuart Hampshire y Donald Davidson sobre el tema de la acción, su enfoque es bastante diferente de la mayor parte de la literatura de la filosofía analítica de los últimos treinta años. En ella se comienza habitualmente con la suposición de que las personas y animales son organismos físicos, preguntándose cómo es posible atribuirles estados psicológicos a tales organismos sobre la base de la observación de su conducta, y sobre la relación entre esos estados psicológicos con los estados cerebrales de los que dependen. Ese enfoque va de afuera hacia adentro. En la mayoría de los casos, los resultados han sido crudos y superficiales, producto de una ceguera autoimpuesta.

El enfoque de O'Shaughnessy va de adentro hacia afuera. El cree, también, que somos (al menos) organismos físicos, pero quiere ver qué podemos descubrir sobre la relación entre nuestros estados psicológicos y la operación de nuestro cuerpo, partiendo del punto de vista del sujeto psicológico, que tiene conciencia en forma inmediata y en ciertos aspectos de lo que está haciendo, sintiendo y pensando. En lugar de tratar de construir la mente a partir de una ontología, desde la cual ha sido excluida, comienza con la mente, la explora desde adentro y descubre que se abre inevitablemente hacia el mundo físico, en virtud de su inevitable posesión de un cuerpo: particularmente en la acción, "la misma alma del agente, expandiéndose más allá de sus confines naturales, hacia afuera, en el mundo en extenso".

O'Shaughnessy describe su teoría de la voluntad como una teoría del doble aspecto, porque sostiene que las acciones voluntarias son esencialmente sucesos tanto psicológicos como físicos —no *combinaciones* de sucesos psicológicos y físicos, sino de dos caras— y que su carácter psicológico requiere que sean también físicos, aunque al mismo tiempo no pueden reducirse o ser analizados en términos de su carácter físico.

Mientras que no trata la relación entre sucesos mentales y el cerebro en general, ni argumenta a favor de una teoría general del aspecto dual de la mente, explica por qué una posición de este tipo puede tener un atractivo contemporáneo:

> Por un lado, los constantes logros de la física y su brillantemente exitosa aplicación en el problema de la vida en Biología Molecular, junto al reciente éxito en la construcción de inteligencias artificiales, crean una línea de pensamiento que conduce al *materialismo.* Por el otro lado, el rechazo creciente al reduccionismo destructivo y nivelador, que está perdiendo rápidamente sus encantos y credibilidad, crean una atmósfera en la que un programa de *conservación ontológica* se vuelve muy atractivo.

"Por supuesto", agrega O'Shaughnessy, "no podemos, al adoptar una perspectiva tan histórica, elevarnos por sobre la necesidad de argumentos." Un sutil motivo antihistoricista.

La investigación de O'Shaughnessy procede con un método de razonamiento a priori, que convierte a su obra en tradicional, en el sentido en que la obra de Kripke es tradicional. Porque se basa en la convicción de que, separando ciertos conceptos naturales e inevitables, por medio de una auto-exploración filosófica que sólo sus poseedores pueden llevar a cabo, podemos hacer descubrimientos fundamentales sobre la naturaleza de la realidad a la que esos conceptos se aplican. Hay muchas cosas que la filosofía sola no puede descubrir, pero tampoco podemos esperar aprender lo más importante de algunos aspectos de la realidad, sin descubrir aquellos rasgos fundamentales con los cuales el pensamiento los capta, y es probable que nos confundamos en nuestra investigación ulterior, si no entendemos primero esos rasgos.

La mente es uno de ellos, y la acción es un ejemplo central. Se puede descubrir mucho sobre ella con métodos empíricos científicos. Pero "los constituyentes últimos de la acción física... son dados en forma precientífica, *a priori.* En este sentido, el concepto de una acción física es un concepto *a priori* que, sin embargo, necesita *a posteriori* condiciones de aplicación dadas". La ciencia puede descubrirlas, por ejemplo los mecanismos fisiológicos que posibilitan la acción, pero la filosofía debe descubrir qué *es* la acción, sus condiciones necesarias y suficientes. Porque, ¿cómo adquirimos esos conceptos a priori?, y ¿por qué deberíamos creer lo que ellos parecen revelarnos acerca de la realidad? La respuesta de O'Shaughnessy es (sin proponérselo) oscura: *"Ciertos hechos naturales deben haber guiado, en forma no argumentativa, a la humanidad prerracional, al formarse estos usos cruciales* en la oscura prehistoria de la espe-

cie, en el amanecer de la autoconciencia. Más exactamente: el pasaje de prerracional a racional involucraba *internamente* la formación de *ciertos hechos naturales* . . . Así es como el hombre pudo conocer *mundo / espacio / verdad / yo / tiempo / conciencia:* toda la galaxia de las Principales Luces Metafísicas". Estos son conceptos que no podrían faltarnos, en tanto seres autoconscientes, racionales. Están determinados por hechos fundamentales de la realidad interna y externa.

En cambio, hay muchos conceptos que podríamos no haber tenido y que en un principio adquirimos ostensiblemente. Casi todo lo concerniente a estas cosas puede ser descubierto por medio de la investigación empírica, porque esos conceptos tienen muy poco contenido.

> Mientras que hay un análisis filosófico del concepto de acción física, casi no podría haber un análisis filosófico comparable de los conceptos idiosincrásicos, puramente *a posteriori*, de oro o de tigre. Pero la divergencia en la naturaleza de estas tareas no es de ningún modo total porque, desde el punto de vista de *la extensión del mundo,* están de acuerdo. Porque el análisis filosófico, desde ese punto de vista, no hace nada más que indicar ciertas restricciones dadas *a priori*, sobre la extensión del "suceso inmediatamente deseado". Le deja a la ciencia la determinación de la extensión del elemento final. Esta última es, precisamente, la posición con "oro" y "tigre".

En cuanto a qué grado de verdad sobre la realidad se esconde en esos conceptos ricos, oscuros y a priori, es un tema controvertido. Muchos filósofos contemporáneos sospechan que esos conceptos naturales están llenos de confusión e ignorancia, de manera que sólo podemos confiar en ellos con cautela, sin dudar en reemplazarlos por otros más claros o más probados científicamente. Creo que esto es, básicamente, erróneo, porque la ciencia tiene que construirse sobre una toma de conciencia precientífica y una comprensión de ciertos aspectos del mundo que nunca pueden ser descartados, sin importar cuánto les agreguemos, porque siempre se los vuelve a usar. Debe ser posible explicar la jerga técnica, en última instancia, en términos naturales; si no, se pierde la conexión con la comprensión, y el peligro de confusión o vacío se vuelve realmente serio.

Algunos pueden encontrar que el método de O'Shaughnessy deposita demasiada confianza en conceptos precientíficos, pero

a mí me parece que está justificado, tanto por sus resultados como por sus premisas.

Estas son algunas de sus conclusiones: la acción física es psicológicamente primitiva y, como la sensación, tiene un carácter independiente de su contenido intencional. No tiene que ser intencional bajo cualquier descripción. Aquí O'Shaughnessy está en desacuerdo con Davidson, y argumenta en forma convincente que cualquier enfoque de la acción voluntaria se debe aplicar a zapateos ociosos y a movimientos de la lengua, de los cuales no necesariamente tenemos conciencia y que no hacemos intencionalmente pero que, sin embargo, no son involuntarios. Estos actos "subintencionales" tienen que contener el núcleo común de toda acción voluntaria. Esta no consiste en un acto mental junto con los movimientos físicos que causa. En cambio, el movimiento físico observable "es la punta superficial de un suceso que se sitúa más atrás, en el cerebro; a saber, el acto de la voluntad". Cada acto, sea exitoso o no, es idéntico al acto de tratar, o esforzarse por, activar una "línea de poder" física preestablecida, desde el cerebro hacia los músculos. Los esfuerzos son la expresión directa e inevitable del acto-deseo: "Son, por así decirlo, nada más que capullos en el árbol del deseo. Y, como cualquier capullo en circunstancias propicias, se convertirán en una flor, de la que nunca serán distintos".

Son sucesos psicológicos porque, al igual que las sensaciones, son accesibles en forma introspectiva. Si creo que estoy tratando de levantar mi brazo, no puedo estar equivocado, aunque mi brazo no se esté levantando. Pero cuando logro levantar mi brazo, mi levantar el brazo es idéntico a mi tratar de levantarlo. Un suceso psicológico tan accesible introspectivamente como este es idéntico a un evento físico externamente observable.

Por lo tanto, llegamos a una teoría del doble aspecto de la acción física. No puede ser reemplazada por una teoría de la volición que desglose la acción en causa psicológica y efecto físico, porque tratar de levantar mi brazo no es un suceso del mundo interno. Cuando logro levantarlo, es una acción observable, y cuando no lo logro, se parece a un comienzo fraccional de esa misma acción, extendiéndose desde su origen en el cerebro hasta el límite de su desarrollo causal, que está bloqueado.

La voluntad en tanto tal no es libre, pero, si resulta de las intenciones de un ser racional cuyos compromisos se encuen-

tran perpetuamente abiertos a revisión o renovación, puede manifestar la libertad posible para aquellos que no son omnipotentes. (No trataré de resumir la compleja teoría causal que relaciona acciones subintencionales, deseo, creencia, decisión, elección, conciencia, intención, compromiso y autodeterminación, y que establece el sentido en el que nuestras acciones no *nos ocurren a nosotros*.)

"Una condición necesaria para un cambio voluntario en un objeto físico es que esté incorporado a la imagen corporal." Pero la imagen corporal no es una entidad sensorial tridimensional, aunque las sensaciones o los datos sensoriales desempeñen un papel en la percepción externa. "Tener la apariencia de tener conciencia de un brazo no es tener conciencia de un brazo aparente." Mejor dicho, la imagen corporal es el *yo corpóreo*, la concepción fundamental de lo que uno es físicamente (y una condición absoluta para conocer la realidad en el modo físico). Hay imágenes corporales de corto y de largo plazo; y pese a que dependen causalmente de la sensación cinestésica y propioceptiva, no son imágenes sensoriales, sino de un tipo mental diferente, que nos permite ubicar nuestras sensaciones corporales y le proporciona a nuestra voluntad algo sobre lo que puede actuar. La imagen corporal de largo plazo "es la que pone en marcha la transición de las sensaciones dadas en un instante, a las posibles posturas y movimientos que uno puede intentar hacer" (aunque no esté limitado a aquello sobre lo que tenemos poder activo: pensar en los pechos de una mujer).

La influencia de Freud sobre este libro es evidente, no sólo en las referencias, sino también en la ambición y esperanza de crear una teoría rica sobre el yo, que sea desde adentro y usando herramientas especializadas, creadas, en gran parte, por él mismo, para descubrir lo que ocurre permanentemente dentro de nosotros. También se nota la influencia de Wittgenstein, una determinación para permitir "extrañas y salvajes manifestaciones del inconsciente filosófico" (esta frase es de O'Shaughnessy) que operan como indicios para descubrir lo que está ante nuestros ojos, en la sutil operación de conceptos que nos presentan con una superficie lisa y simple.

El resultado es una teoría de la mente y del cuerpo mucho más rica de lo que Wittgenstein se hubiera permitido tener. Pero comparte con las *Investigaciones filosóficas* una posición

según la cual, si bien los fenómenos mentales son sui géneris y no analizables, ni en términos de la conducta ni en alguna otra cosa respecto del cuerpo, están esencialmente conectados con el cuerpo y entre sí. La manera de acercarnos lo más posible a una definición de estos conceptos es descubrir sus conexiones esenciales. Y esto no es lo mismo que reducir los contenidos de la mente a espacios en blanco dentro de la red de circunstancias corporales y manifestaciones, con los cuales se encuentran epistémicamente o causalmente relacionados. Un dolor de cabeza es tan real como un golpe en la cabeza.

10

Searle: ¿Por qué no somos ordenadores?*

Este ensayo fue originalmente una reseña de The Rediscovery of the Mind *(M.I.T. Press, 1992), de John Searle. Aunque Searle y yo estamos de acuerdo en muchas cosas, no creo que sea posible distinguir su solución antirreduccionista del dualismo de propiedad. Tampoco creo que pueda ser un hecho bruto de la naturaleza que las propiedades de orden superior del sistema nervioso sean producidas por las particularidades de su operación fisioquímica. La relación entre los niveles tiene que ser más "interna" que eso, es decir, una forma de consecuencia inteligiblemente necesaria, antes que una mera correlación. La irreductibilidad de lo ontológicamente subjetivo a lo ontológicamente objetivo sigue siendo un obstáculo a la imaginabilidad de tal conexión.*

Según una posición ampliamente aceptada, el cerebro es un ordenador gigante, y la relación entre la mente y el cerebro humanos es semejante a la relación entre el programa del ordenador y el hardware electrónico sobre el que el programa funciona. El filósofo John Searle, un luchador por temperamento, se propuso mostrar que esta posición y la tradición materialista subyacente en ella, son un sinsentido por diferentes razones, algunas de las cuales son obvias y otras más sutiles. Mediante la elaboración de argumentos que él y otros sostuvieron en los últimos veinte años, Searle ataca la mayor parte

* Reimpreso con autorización de *The New York Review of Books*, copyright ©1993 por Nyrev, Inc. Publicado por primera vez el 4 de marzo de 1993.

de la estructura de la ciencia cognitiva, y luego ofrece una teoría propia sobre la naturaleza de la mente y su relación con el mundo físico. Si este libro punzante está en lo cierto, el modelo computacional de la mente es, no sólo dudoso o imperfecto, sino total y obviamente absurdo.

Las principales razones son dos. Primero, la esencia de la mente es la conciencia: todos los fenómenos mentales son real o potencialmente conscientes. Y ninguno de los análisis materialistas familiares de la mente pueden tratar con la experiencia consciente: la dejan afuera, porque no hablan de ella, o porque la identifican con algo que no tiene nada que ver con ella. En segundo lugar, los ordenadores, que no tienen mente, pueden ser descritos como funcionamiento de programas, procesamiento de información, manipulación de símbolos, respuestas a preguntas, etc., sólo porque están construidos para que la gente, que sí tiene mente, pueda interpretar sus operaciones físicas de ese modo. Atribuirle al cerebro un programa de ordenador implica una mente que pueda interpretar lo que el cerebro hace, de modo que la propuesta de explicar la mente en términos de dicho programa es incoherente.

I

El libro de Searle comienza con un lúcido repaso crítico de las diferentes posiciones actuales sobre la relación entre la mente y el cuerpo. El problema mente-cuerpo fue formulado en su versión moderna recién en el siglo XVII, cuando surge la concepción científica del mundo físico con la que todos nosotros nos formamos actualmente. Según esta concepción, el mundo físico es, en sí mismo, incoloro, inodoro e insípido, y puede ser descrito matemáticamente por leyes que gobiernan la conducta de partículas y campos de fuerza en términos de espacio y tiempo. Algunos fenómenos físicos causan en nosotros experiencia perceptiva —vemos colores y oímos sonidos— pero las cualidades que experimentamos no pertenecen a las ondas de la luz y del sonido descritos por la física. Llegamos a la realidad física por medio de una "depuración" de los efectos subjetivos en nuestros sentidos y de la forma en que las cosas aparecen desde el punto de vista humano, atribuyéndoselas a la mente, y tratando de construir una teoría objetiva del mundo

fuera de nuestra mente, que explique sistemáticamente las observaciones y mediciones experimentales sobre las que todos los observadores escrupulosos están de acuerdo. Por más que el contenido de la física contemporánea y su concepción sobre el papel del observador se diferencie radicalmente del contenido de la física clásica, sigue buscando una teoría del mundo *externo*, en este sentido.

Pero, una vez que se alcanza dicha concepción, al eliminar las apariencias del mundo físico y asignárselas a la mente, la ciencia debe enfrentar el problema de cómo completar la imagen, encontrando un lugar en el mundo para nuestras mentes, con sus experiencias perceptivas, pensamientos, deseos, construcción de teorías científicas, y muchas otras cosas que la física no describe. La razón por la que esto se llama problema mente-*cuerpo* es que lo que ocurre en nuestra mente evidentemente depende de lo que ocurre a y en nuestro cuerpo, especialmente a nuestro cerebro. Sin embargo, nuestro cuerpo es parte del mundo "externo", es decir, el mundo externo a nuestra mente, descrito por la ciencia física. Nuestro cuerpo es una estructura física elaborada, compuesta por moléculas, y la física y la química presumiblemente deberían hacer una descripción muy precisa de todo lo que aquel hace y experimenta.

Es sabido que Descartes pensó que, si consideramos cuidadosamente la naturaleza de la realidad física externa y la naturaleza mental interna (ejemplificada por nuestra propia mente), no podemos dejar de notar que ambas tienen que ser dos tipos diferentes de cosas, independientemente de cuán unidas estén: una mente y sus pensamientos y experiencias no sólo puede estar constituida por partes físicas, como moléculas, que es como evidentemente están constituidos nuestro corazón y cerebro. La conclusión de Descartes de que la vida mental ocurre en una entidad no física, a saber, el alma, es conocida como dualismo: a veces, como dualismo "sustantivo", para distinguirlo del dualismo de "propiedad", que considera que, aunque no existe un alma distinta del cuerpo, los fenómenos mentales (como el gusto salado, o tener sed) suponen propiedades de la persona o de su cerebro que no son físicas.

El poder del argumento intuitivo de Descartes es considerable pero, hoy en día, ambas clases de dualismo tienen pocos adeptos entre los filósofos,[1] la mayoría de los cuales aceptan algún tipo de materialismo. La mayor parte de los filósofos cree

que todo lo que existe y todo lo que ocurre en el mundo debería poder ser descrito por la ciencia física. Más aún, encuentran evidencia directa de que esto es posible aun para la mente, en la íntima dependencia de lo mental de los procesos neurofisiológicos, sobre los que se ha aprendido mucho desde el siglo XVII hasta hoy. Y encuentran evidencia directa a partir de la exitosa aplicación de la física y la química a otros aspectos de la vida, desde la digestión hasta la herencia. Consecuentemente, la mayoría de los esfuerzos para completar la cosmovisión científica en forma materialista han llevado a cabo una especie de reducción de lo mental a lo físico, donde lo físico, por definición, es lo que puede ser descrito en términos no mentales.

Una reducción es el análisis de algo identificado en un nivel de descripción, en términos de otro nivel de descripción considerado más fundamental, que nos permite decir que el primero en realidad no es sino el segundo: el agua puede describirse como moléculas de H_2O, el calor como movimiento molecular, la luz como radiación electromagnética. Estas son reducciones de lo macroscópico físico a lo microscópico físico, y tienen los siguientes rasgos notables: (1) proveen no sólo información externa sobre las causas o condiciones del fenómeno reducido, sin una explicación interna de lo que realmente es el agua, el calor o la luz; (2) funcionan sólo porque hemos distinguido las apariencias perceptivas de los fenómenos macroscópicos —cómo sentimos el agua y el calor, cómo vemos la luz— de las propiedades que estamos reduciendo. Cuando decimos que el calor consiste en movimiento molecular, queremos decir que el calor, como la propiedad intrínseca de objetos calientes, no es sino el movimiento de sus moléculas. Tales objetos producen en nosotros la sensación de calor cuando los tocamos, pero no identificamos esa sensación con el movimiento molecular; en efecto, la reducción depende de que lo hayamos dejado de lado.

Ahora bien, ¿cómo podríamos reducir los fenómenos *mentales* a algo descrito totalmente en términos físicos, no mentales? En este caso, no podemos dejar de lado todos los efectos sobre la mente, ya que eso es precisamente lo que tenemos que reducir. Lo que necesitamos para completar la imagen del mundo materialista es algún esquema del tipo: "los fenómenos mentales (pensamientos, sentimientos, sensaciones, deseos, percepciones, etc.) no son sino...", donde tenemos que completar la

frase con una descripción que sea explícitamente física, o que utilice sólo términos aplicables a algo totalmente físico.[2] Los diferentes intentos de llevar a cabo esta tarea, en apariencia imposible, constituyen la historia de la filosofía de la mente de los últimos cincuenta años.

La participación de Searle en esta empresa parte del conductismo, la concepción de que los conceptos mentales no se refieren a nada dentro de nosotros, y que cada tipo de estado mental puede ser identificado con una disposición del organismo para conducirse de determinadas maneras bajo ciertas condiciones físicas. Cuando esta posición parecía convertirse en un franco rechazo a la existencia de la mente, algunos filósofos propusieron teorías de la *identidad*, según las cuales los procesos mentales son idénticos a los procesos cerebrales, del mismo modo que la luz es idéntica a la radiación electromagnética. Pero las teorías de la identidad tenían el problema de explicar en términos no mentales, qué quiere decir de un proceso cerebral determinado, que es un pensamiento o una sensación. Después de todo, no quiere decir solamente que es cierto tipo de proceso neurofisiológico. Dado el objetivo de estas teorías, no podía querer decir que el proceso cerebral tiene algún efecto mental. La solución propuesta era un renacimiento del conductismo, con una nueva forma: la sed, por ejemplo, se identificaba no sólo con una disposición a beber, sino con un estado cerebral, pero ese estado cerebral determinado, al ser idéntico a la sed, ahora consistía simplemente en el hecho de que era típicamente causado por deshidratación, y que causaba típicamente la disposición a beber. De este modo, se pensaba que la identificación de estados mentales con estados cerebrales podía evitar toda referencia a rasgos no físicos.

A estos análisis "conductistas causales" les sucedió, eventualmente, una teoría más técnica, llamada "funcionalismo", según la cual los conceptos mentales no pueden estar unidos individualmente a la conducta y circunstancias, sino como una parte de una única red interconectada. La conducta causada por la sed, por ejemplo, depende del resto de la condición mental de la persona, es decir, sus creencias sobre dónde encontrar agua, y sobre si el agua es potable, la fuerza de sus deseos de vivir o morir, etc. Cada estado mental forma parte de un sistema integrado, que controla la interacción del organismo con el

medio: sólo mediante el análisis del papel que desempeñan los estados como sed, dolor, otros tipos de sensación, creencia, emoción y deseo, dentro del sistema en su totalidad, se puede describir adecuadamente su conexión con la conducta y las circunstancias externas. Un sistema de este tipo permite identificar estados mentales con estados cerebrales, siempre que los últimos tengan roles causales o funcionales del tipo especificado por la teoría (que debe ser construida) de cómo funciona el sistema integrado. Finalmente, el funcionalismo condujo a lo que Searle denomina IA Fuerte (Inteligencia Artificial Fuerte), a saber, la identificación de estados mentales con estados computacionales de un programa de ordenador que controla la conducta del organismo, un programa que se lleva a cabo físicamente en el hardware (o wetware) del cerebro.[3]

Todas estas teorías intentan reducir la mente a algún aspecto de un mundo que pueda describirse completamente por la física, es decir, el mundo de las partículas y campos. Estas teorías no fueron desarrolladas en detalle, sólo son esperanzadas descripciones de lo que debería ser una teoría de la mente, junto con algunos ejemplos extremadamente esquemáticos. En tanto que cada nueva propuesta fue criticada por defensores de algún enfoque reduccionista alternativo, Searle argumenta que todas ellas tienen un grave defecto: todas dejan a la conciencia de lado.

II

Ninguna teoría que deje la conciencia de lado puede pretender ser una teoría de la mente, y no es posible hacer un análisis de la conciencia en términos no mentales. Consecuentemente, ninguna reducción materialista de lo mental puede tener éxito. Searle arguye que ninguna de estas teorías podría decir lo que realmente es dolor, hambre, creencia, visión, etc., porque todo lo que tratan es externamente observable —la conducta del organismo y su estructura causal—, y una descripción exclusivamente en esos términos no puede garantizar que tiene conciencia: cada una de estas teorías conductistas, funcionalistas o computacionales podrían satisfacerse por un autómata inconsciente dotado de la suficiente complejidad física.

> La pregunta crucial no es "¿Bajo qué condiciones *atribuiríamos* estados mentales a otras personas?, sino, "¿Qué es lo que *realmente tiene* la gente cuando tiene estados mentales?", "¿Qué son los fenómenos mentales?", en tanto distinta de "¿Cómo conocerlos, y cómo funcionan causalmente en la vida del organismo?". (p.23)

Les atribuimos conciencia a otras personas y a los animales, sobre la base de su conducta, lo cual es simplemente una evidencia de conciencia, más bien que una prueba de conciencia, y debe ser complementada con evidencia de similaridad fisiológica: ya que creemos en la uniformidad de la naturaleza, inferimos naturalmente que criaturas que se conducen en forma similar a nosotros, y que tienen órganos sensoriales y sistemas nerviosos similares físicamente a los nuestros, también tienen experiencias conscientes de alguna clase. Pero, Searle argumenta, ninguna cantidad de hechos sobre la conducta física o la organización funcional implica por sí misma que un sistema sea consciente. Y, una teoría que sostiene, por ejemplo, que la visión no es "nada más que" un estado del organismo, a fin de ser adecuada, debe tener la consecuencia de que, si el organismo está en ese estado, *no puede* dejar de ser consciente. De lo contrario, dejará de lado lo más importante sobre la visión y, por más ventajas que ofrezca, no podrá representar un enfoque válido sobre lo que es la visión.

Los reduccionismos materialistas no sólo no pueden implicar que el sistema es consciente; es evidente de antemano que ningún desarrollo posterior en esta misma dirección, ni ninguna complicación adicional estructural o conductista, podrían hacerlo. La razón es que hay una diferencia crucial entre fenómenos conscientes y fenómenos de conducta o fisiológicos, que hace que los primeros sean irreductibles a los segundos: la conciencia es, en términos de Searle, "ontológicamente subjetiva". Es decir, sus rasgos esenciales no pueden ser descritos totalmente desde un punto de vista externo de una tercera persona. Incluso la descripción fisiológica de lo que ocurre dentro del cráneo es, en este sentido, externa: la describimos desde afuera. No basta con resumir las observaciones, conductuales o fisiológicas, de una tercera persona, que nos permiten atribuirles estados mentales conscientes a los demás. El punto de vista de la primera persona, que nos revela cómo es un estado mental consciente para el sujeto, es indispensable.

Esto se torna evidente cuando preguntamos: ¿qué es la conciencia? Aunque podamos describir ciertos rasgos e identificar más tipos específicos de fenómenos mentales, en tanto sus instancias, es tan básica que no puede ser definida en términos de otra cosa. Usted, lector, es consciente, en este preciso momento, y su condición de conciencia incluye cosas como su manera de percibir esta página, cómo siente el papel entre sus dedos, cómo siente su camisa en el cuerpo y la silla sobre la que está sentado, los sonidos que está escuchando, la música, o las sirenas de la policía en el fondo; y su experiencia de leer esta oración. La posición de Searle es que el análisis de la tercera persona, sea conductual, causal o funcional, no importa cuán extenso sea, no puede decirnos qué son estas experiencias en sí mismas, en qué consisten, en tanto distintas de sus causas y efectos. Esto es absolutamente obvio, porque los hechos subjetivos sobre lo que es para una persona estar en ciertas condiciones no pueden ser idénticos a los hechos sobre lo que las cosas son, ni desde el punto de vista de una persona, ni *para* nadie, sino en sí mismos. Los hechos sobre nuestra conducta externa o la actividad eléctrica u organización funcional de nuestro cerebro pueden estar conectados de cerca con nuestra experiencia consciente, pero no son hechos sobre lo que es para mí escuchar una sirena de policía.[4]

Searle cree que la insistencia en el reduccionismo materialista, a pesar de su evidente falsedad, requiere una explicación. El la compara con la repetición constante de una conducta destructiva por parte de un neurótico compulsivo, y espera, al traer a la luz sus causas subyacentes, poder romper con esa compulsión. Es evidente, tanto por lo que dicen como por lo que hacen, que los reduccionistas están convencidos de antemano de que alguna teoría materialista debe ser correcta: lo que tienen que hacer es encontrarla. Esta suposición es parte de una cosmovisión científica que, según ellos, no tiene alternativa. Pero, según Searle, hay dos errores conceptuales cruciales que subyacen en esta suposición. El primero de ellos es que debemos elegir entre materialismo y dualismo:

> No quiero dejar de insistir en que uno puede aceptar los hechos obvios de la física —por ejemplo, que el mundo consiste enteramente en partículas físicas, en campos de fuerza— sin negar al mismo tiempo los hechos obvios de nuestras propias experiencias, por ejemplo, que somos todos conscientes y que nuestros estados

> conscientes tienen propiedades fenomenológicas *irreductibles* [es decir, subjetivas] bastante específicas. El error consiste en suponer que estas dos tesis son incompatibles, y que el error proviene de aceptar los supuestos del vocabulario tradicional. Mi posición no es, de ningún modo, una forma de dualismo. Rechazo tanto el dualismo de propiedad como el sustantivo. Pero, precisamente por las razones por las que rechazo el dualismo, rechazo el monismo y también el materialismo. El error profundo es suponer que debemos optar entre estas posiciones . . .
> Una vez que aceptamos nuestra cosmovisión, el único obstáculo para garantizarle a la conciencia su carácter de rasgo biológico de un organismo, es la suposición, pasada de moda, dualista/materialista, de que el carácter "mental" de la conciencia le impide ser una propiedad "física". (pp. 28, 91)

Esta tesis radical de que la conciencia es una propiedad física del cerebro, a pesar de su subjetividad, y de que es irreductible a cualquier *otra* propiedad física, es el corazón metafísico de la posición de Searle. Sin embargo, lo importante es que Searle sostiene que los materialistas se dejan conducir hacia formas imposibles de reducción psicofísica, porque suponen que, si los estados mentales no pueden explicarse en tales términos, entonces la alternativa ineludible es el dualismo: en consecuencia, deberían admitir que las sustancias o propiedades no físicas son rasgos básicos de la realidad. Y el temor al dualismo, con sus asociaciones religiosas, espiritualistas y no científicas, los lleva a adoptar el materialismo reduccionista a cualquier precio: "Entonces, el materialismo es, en cierto sentido, la flor más fina del dualismo".

Para escapar de este brete, según Searle, tenemos que liberarnos de la urgencia de preguntarnos si hay una o dos clases últimas de cosas y propiedades. En primer lugar, no deberíamos empezar por contarlas.

Searle está absolutamente en lo cierto sobre el temor al dualismo (en efecto, creo que él mismo no es inmune a él). Su expresión más extraña es, sin embargo, otra teoría llamada "materialismo eliminativo". Se trata de la posición de que, dado que los estados mentales no pueden acomodarse dentro del mundo descrito por la física, no existen, así como no existen ni las brujas ni los fantasmas. Pueden ser rechazados como postulados de una teoría primitiva, a la que habitualmente nos referimos como "folk psychology". Sobre esta teoría, P. F. Straw-

son ha dicho que es la provincia de "gente tan sencilla como Shakespeare, Tolstoy, Proust y Henry James".[5] Pacientemente, Searle pulveriza esta posición, pero el núcleo de su argumentación es que toda la tradición materialista es eliminativa: todas las teorías materialistas niegan la realidad de la mente, pero la mayoría de ellas disfraza el hecho (por sí mismos o por los demás), al identificar la mente con alguna otra cosa.

El segundo error conceptual crucial que trasciende la búsqueda compulsiva de teorías materialistas, según Searle, es el error, simple pero enormemente destructivo, sobre la objetividad:

> Hay una constante confusión entre la afirmación de que deberíamos tratar, en lo posible, de eliminar los prejuicios subjetivos personales en la búsqueda de la verdad, y la afirmación de que el mundo real no contiene ningún elemento que sea irreductiblemente subjetivo. Esta confusión, a su vez, se basa en una confusión entre el sentido epistémico y el sentido ontológico de la distinción subjetivo/objetivo. En el sentido epistémico, la distinción marca diferentes grados de independencia de afirmaciones, desde los caprichos de valores especiales, prejuicios personales, puntos de vista y emociones. En el sentido ontológico, la distinción marca diferentes categorías de realidad empírica. (p. 19)

Esto me parece totalmente convincente, y muy importante. Naturalmente, la ciencia debe buscar objetividad epistémica —conocimiento objetivo— usando métodos que compensen los diferentes puntos de vista y que permitan a diferentes observadores llegar a la misma concepción de algo. Pero es una gran confusión concluir, a partir de eso, que nada que tenga o incluya un punto de vista pueda ser objeto de investigación científica. Los puntos de vista subjetivos son, ellos mismos, partes del mundo real y, si ellos y sus propiedades deben ser descritos adecuadamente, su carácter ontológicamente subjetivo —la subjetividad de su *naturaleza*— debe ser reconocida. Más aún, es posible hacerlo objetivamente en un sentido epistémico: aunque sólo usted esté teniendo la experiencia de mirar la página ante sus ojos, otros pueden saberlo y pueden saber bastante bien cómo es esa experiencia para usted. Es una verdad objetiva que usted está teniendo cierta experiencia visual subjetiva en este momento.

Si aceptamos esta distinción, la pregunta es: ¿cómo podemos formarnos una concepción científica objetiva, desde el punto de vista epistémico, de un mundo que no contiene solamente los hechos objetivos ontológicamente familiares, descritos por la física, química, biología, sino también los hechos ontológicamente subjetivos de la conciencia? Y esta pregunta nos trae, finalmente, a la posición de Searle, que denomina "naturalismo biológico", y que combina la aceptación de la irreductible subjetividad de lo mental con el rechazo a la dicotomía entre lo mental y lo físico: "la conciencia... es un rasgo biológico del cerebro humano y de los de ciertos animales. Está causada por procesos neurobiológicos y es tanto parte del orden biológico natural como cualquier otro rasgo biológico, ya sea fotosíntesis, digestión o mitosis". (p. 90)

Y, a pesar de su antirreduccionismo, también dice lo siguiente:

> La conciencia es una propiedad de nivel superior, o emergente, del cerebro, en el sentido totalmente inocuo de "nivel superior" o "emergente", en el cual la solidez es una propiedad emergente de nivel superior de moléculas de H_2O, cuando se encuentran en una estructura entrelazada (hielo), y la liquidez es, análogamente, una propiedad emergente de nivel superior de moléculas de H_2O cuando están, por así decirlo, rodando unas con otras (agua). La conciencia es una propiedad mental, por lo tanto física, del cerebro, en el sentido en que la liquidez es una propiedad de sistemas moleculares. (p.14)

Si pudiéramos aclarar esta posición, como para distinguirla de las alternativas, debería representar un aporte fundamental a las respuestas posibles al problema mente-cuerpo. Pero no creo que lo sea.

Supongamos que garantizamos que los estados de la conciencia son propiedades del cerebro causadas por, pero no reducibles a, su actividad neuronal. Esto quiere decir que mi cerebro, por ejemplo, tiene un punto de vista, del que mis experiencias actuales son aspectos. Pero, ¿cuál es la justificación para llamar *físicos* a estos rasgos del cerebro, que son irreductibles y subjetivos? ¿Qué significa llamarlos físicos? Por cierto, son de "nivel superior" en el sentido de que pueden ser atribuidos sólo al sistema como un todo y no a sus partes microscópicas; también son "emergentes" en el sentido de que se explican

sólo por las interacciones causales de aquellas partes. Pero, por mayor que sea la variedad de fenómenos físicos, la subjetividad ontológica es una de sus condiciones definitorias centrales. Asimismo, como ya vimos, Searle insiste en que la conciencia es ontológicamente subjetiva.

Searle no dice demasiado sobre este punto. Quizá cree que, si los cerebros están constituidos por partículas físicas, se deduce automáticamente que todas sus propiedades son físicas. Y cita una observación de Noam Chomsky, quien sostiene que, en cuanto comprendemos algo, lo llamamos "físico". Pero, si "físico" es, en este sentido, un término meramente honorífico (este es otro término que Chomsky ha usado), entonces, ¿cuál es el contenido metafísico de la afirmación de Searle de que las propiedades mentales son físicas, y su enfático rechazo al dualismo de propiedad? El dice, después de todo, que la distinción ontológica entre subjetivo y objetivo señala "diferentes categorías de la realidad empírica". Pero decir, además, que nos "quedamos con un universo que contiene un componente físico irreductiblemente subjetivo como componente de la realidad física", sólo conlleva una postura esencialmente dualista en el lenguaje, que expresa una fuerte aversión al dualismo.[6]

Quizá deberíamos adoptar el uso de Searle de la palabra "físico", pero el problema central es más que un problema verbal. Es el problema de cómo construir una cosmovisión científica inteligible y completa, una vez que rechazamos la reductibilidad de lo mental a lo no mental. Tal como señala Searle, para lograr este objetivo, no podemos seguir por el camino que la ciencia ha seguido desde el siglo XVII, porque este dependía de la exclusión de la mente del observador del mundo que se observa y describe. Proponer a la conciencia como una propiedad subjetiva intrínseca del cerebro, causada por su actividad neuronal, es el primer paso en un camino diferente: el camino correcto, en mi opinión. Pero hay grandes problemas a resolver, y no se trata sólo de problemas empíricos, sino también filosóficos.

Aun si aprendemos mucho más de lo que sabemos hoy en día sobre las causas fisiológicas de la conciencia, no hará, como Searle ya sabe, a la relación de la conciencia con la conducta de neuronas análoga a la relación de la liquidez con la conducta de moléculas de H_2O. En este último caso, la relación es transparente: podemos *ver* cómo la liquidez es el resultado lógico de las moléculas "rodando unas con otras" en el nivel microscópi-

co. No podemos esperar nada comparable en el caso de las neuronas, aun cuando sea empíricamente evidente que los estados de conciencia son las consecuencias necesarias de la actividad neuronal. Searle hace un análisis interesante sobre esta diferencia que, dice, resulta no sólo de una limitación de nuestros poderes para concebirlo: podemos representarnos la relación necesaria entre los niveles macro y micro del agua, ya que nos los imaginamos desde afuera, pero no podemos hacerlo con la subjetividad, que nos tenemos que imaginar desde adentro, sea la nuestra o la de los demás. Estoy de acuerdo, pero creo que esto significa que no entendemos realmente la afirmación de que los estados mentales son estados del cerebro: todavía somos incapaces de concebir *cómo* la conciencia surge en la materia, incluso si estamos seguros de que lo hace.[7]

III

La segunda serie de argumentos de Searle en contra del modelo computacional de la mente depende de la naturaleza específica de los ordenadores, y es más típicamente propia de Searle: surge de su antigua preocupación por la teoría del significado y la "intencionalidad" de los estados mentales: es decir, su capacidad para significar algo o referirse a algo fuera de ellos mismos, y su consecuente susceptibilidad respecto de los juicios de verdad o falsedad, y corrección o incorrección.[8]

¿Cómo pueden los ordenadores responder a las preguntas que les hacemos, por ejemplo, las preguntas aritméticas? La explicación tiene dos partes. En primer lugar, es posible formular cada una de las preguntas usando una secuencia de símbolos —letras o numerales— que se seleccionan de una corta lista y se distinguen por sus formas, y diseñar una serie finita de reglas para manipular esos símbolos, la cual tiene la siguiente propiedad: si comenzamos con la secuencia que corresponde a la pregunta, y seguimos las reglas para mover, eliminar y agregar símbolos, llegaremos, después de una serie finita de pasos determinados uno tras otro, a un punto en el que las reglas nos dirán que debemos detenernos; la última secuencia que produjimos corresponderá a la respuesta,[9] que simplemente debemos leer. Pero, a fin de seguir las reglas para manipular los símbo-

los, no necesitamos saber qué significan, si es que significan algo: simplemente, debemos identificar sus formas.

Estas reglas se llaman reglas de la *sintaxis* (en oposición a las reglas de la *semántica*, que necesitamos para interpretar una secuencia que signifique algo). Lo bueno de esto es que podríamos entrenar a alguien en división, por ejemplo, en forma totalmente mecánica, usando una clase de reglas y tablas, sin que él sepa que los símbolos que está escribiendo representan números o alguna otra cosa: si esa persona sigue las reglas sintácticas, llegará a lo que nosotros leeríamos (aunque él no) como la respuesta.[10]

La segunda parte de la explicación es que hay, además de la escritura sobre papel, diferentes maneras de codificar la aritmética en un conjunto de símbolos y reglas sintácticas, y, para alguna de esas maneras, es posible diseñar máquinas físicas a fin de que puedan llevar a cabo mecánicamente todos los pasos en la manipulación de los símbolos que prescriben las reglas. En lugar de la persona que sigue las reglas sintácticas mecánicamente, sin saber lo que significan los símbolos, un mecanismo físico puede hacer las mismas operaciones de los símbolos en forma automática. Pero, nótese que este mecanismo no sólo no sabe lo que los símbolos significan, sino que ni siquiera sabe, a diferencia del escriba de nuestro ejemplo, que ignora la semántica, que está siguiendo reglas para su manipulación, reglas sintácticas. No sabe nada. En realidad, no está siguiendo reglas, sino que está funcionando solamente de acuerdo con las leyes de la física, en las formas diseñadas por ingenieros expertos, que nos permiten interpretar los resultados tanto sintáctica como semánticamente.

El conocido argumento de Searle de la "habitación china" describe una persona consciente que, sin saber chino, sigue reglas de manipulación de los caracteres chinos y arma oraciones inteligibles para personas que saben chino. El punto crucial en contra del modelo computacional de la mente es que la sintaxis sola no puede llegar a la semántica.[11] En *The Rediscovery of the Mind*, Searle extiende el argumento para mostrar que la física sola no puede llegar a la sintaxis. Seguir reglas, aunque sean reglas puramente sintácticas, es un proceso mental irreductible, un proceso "intencional", en el que el significado de las reglas en sí mismas debe ser captado por una mente consciente. No se trata sólo de una cuestión de regularidad en la conducta física.

> Un estado físico de un sistema es un estado computacional sólo cuando se le asigna a ese estado algún papel, función o interpretación computacional . . . *nociones tales como computación, algoritmo y programa no nombran rasgos físicos intrínsecos de sistemas:* no *descubrimos* estados computacionales dentro de la física, se los *asignamos* a la física . . .
> *El objetivo de la ciencia natural es descubrir y caracterizar rasgos que son intrínsecos del mundo natural. Por la misma definición de computación y conocimiento, no hay ninguna forma en que la ciencia cognitiva computacional pueda ser una ciencia natural, porque la computación no es un rasgo intrínseco del mundo. Es asignado en forma relativa a los observadores.* (pp. 210, 212; bastardilla en el original)

La distinción de Searle entre lo que es intrínseco a la cosa observada y lo que es relativo a un observador o intérprete es fundamental. El argumenta que la intencionalidad intrínseca —es decir, la capacidad de captar el significado de enunciados, y seguir reglas conscientemente— ocurre sólo en la mente. Podemos decir que palabras en una página o resistencias eléctricas en un chip de ordenador significan algo o que obedecen reglas de gramática o de aritmética, sólo en el sentido derivativo de que nuestra mente puede interpretarlos de ese modo, en virtud de su disposición. Esto significa que la afirmación de que el cerebro es un ordenador implica que el cerebro tiene intencionalidad y sigue reglas de computación que no son intrínsecas, sino sólo relativas a la interpretación de su usuario. ¿Y quién se supone que es ese usuario? Si el cerebro es un ordenador, no tiene intencionalidad intrínseca. Si la tiene, deber ser algo más que un ordenador. Searle elige esta segunda alternativa. También discute que las teorías que tratan de representar la mente como un ordenador, subrepticiamente suponen una mente o un "homúnculo" que la interprete.

Hay mucho más todavía en este argumento y, aunque su conclusión negativa me parece convincente, se podrían plantear algunas preguntas que no he tratado aquí con respecto a la teoría positiva de Searle de la intencionalidad intrínseca, que, supuestamente, es compatible con su naturalismo biológico. Al igual que la conciencia, es muy difícil ver cómo la intencionalidad intrínseca podría ser una propiedad del organismo físico. Pero, en lugar de ahondar en estas cuestiones, consideraré la posición de Searle sobre el inconsciente y su relación con la conciencia, ya que ellos unen ambas partes del argumento.

Searle le ha dado gran importancia a la afirmación de que la conciencia subjetiva no es reductible a ninguna otra cosa. Pero la mayoría de nuestros estados mentales no son conscientes. Tomemos todas las creencias e intenciones que tenemos, pero en las que no estamos pensando en este momento. Por ejemplo, la creencia de que hay una torre inclinada en Pisa. Recién se ha vuelto consciente, pero probablemente tenemos esta creencia desde hace años. Si tales creencias pueden existir en forma inconsciente, entonces la conciencia no es un rasgo esencial de la vida mental, y debe ser posible que los estados mentales estén incorporados en un cerebro puramente material. De modo que se podría decir que, incluso para aquellos estados mentales que son conscientes, su carácter subjetivo, experimentado, no es esencial para su intencionalidad. Quizá la conciencia sea sólo un tipo de "tinte" subjetivo que algunas veces se les agrega al blanco y negro verdaderamente funcional de los estados mentales, de modo que una teoría de la mente podría dejar de lado a la conciencia, en tanto algo no esencial.

Esta es la respuesta de Searle a esta sugerencia:

> Entendemos la noción de estado mental inconsciente solamente como un contenido posible de la conciencia, sólo como el tipo de cosa que, aunque no es consciente y es quizás imposible de traerse a la conciencia, por varias razones [tales como la represión] es, sin embargo, la *clase de cosa* que puede ser o podría haber sido consciente. (p. 156)

Esto es lo que Searle llama el "principio de conexión"; sus argumentos a favor de él es que incluso los estados mentales inconscientes deben tener un carácter distintivamente subjetivo, que puedan contar como creencias, pensamientos, deseos, etc. Para poder ser sobre algo, y por lo tanto, ser verdadero o falso, correcto o incorrecto, un estado debe pertenecer, al menos potencialmente, al punto de vista de algún sujeto. Searle reconoce que era un estado neurofisiológico de mi cerebro lo que hizo verdadero, hace dos horas, que yo creyera que había una torre inclinada en Pisa, pero argumenta que la neurofisiología sola no puede elevar ese estado al rango de creencia, porque ninguna descripción fisiológica por sí misma implica que el estado cerebral tenga alguna intencionalidad o significado, aun si le agregamos a la descripción la conducta física que podría causar. Su conclusión es que se les puede atribuir inten-

cionalidad a los estados neurofisiológicos sólo en forma derivativa, en virtud de los estados conscientes que pueden producir: "La ontología del inconsciente consiste en rasgos objetivos del cerebro, capaces de causar pensamientos conscientes subjetivos". Esto último tiene la sorprendente consecuencia de que un mecanismo profundo, supuestamente psicológico, como el Instrumento de adquisición del lenguaje de Chomsky, que le permite al niño aprender la gramática de un lenguaje a partir de los segmentos de habla con los que se encuentra en una edad temprana, no es una clase de reglas mentales inconscientes, sino simplemente un mecanismo físico, porque es incapaz de dar origen a un pensamiento consciente subjetivo, cuyo contenido consiste en esas mismas reglas. De manera que, en la posición de Searle, la conformidad del niño a las reglas, al aprender el lenguaje, no es un ejemplo de intencionalidad intrínseca, sino sólo una intencionalidad que le atribuye el observador-lingüista.

En suma, la conciencia es la esencia de lo mental, incluso si la mayoría de los estados mentales no son conscientes en algún momento determinado. No podemos estudiar ningún aspecto de la experiencia mental sin incluirla, o ella o a su posibilidad, en la definición de lo que estamos tratando de entender. En particular, la intencionalidad es inseparable de ella.

The Rediscovery of the Mind es incisivo, agresivo y bellamente claro, en el mejor estilo de Searle de "¿Qué es todo este sinsentido?". Es un valioso antídoto contra una de las ilusiones predominantes de nuestro tiempo.

Notas

1. Véase Geoffrey Madell, *Mind and Materialism* (Edinburgh University Press, 1988).

2. Otra estrategia reduccionista, que no puedo analizar aquí en extenso, consiste en sustituir una teoría de lo que *son* los estados mentales, por una teoría de las bases observables externamente por las que *atribuimos* estados mentales a las personas, y sostener que este sistema de "condiciones de asertividad" es todo el análisis que estos conceptos requieren. No *identificamos* los fenómenos mentales con algo físico, porque no necesitamos identificarlos con nada. Pero las condiciones de aplicación de los conceptos mentales son, según esta posición, compatibles con el hecho de que el mundo es nada más que un sistema material. Esta es fundamentalmente la estrategia de Daniel Dennett en *Consciousness Explained* (Little, Brown, 1991) [analizada en el cap. 8 de este volumen].

3. El conductismo está más o menos representado por Gilbert Ryle, la teoría de la identidad por J. J. C. Smart, y el funcionalismo por Hilary Putnam, pero hay muchos autores, y la bibliografía es muy extensa. Véanse algunas de las excelentes colecciones recientes sobre la filosofía de la mente: Ned Block (comp.), *Readings in Philosophy of Psychology,* 2 vols. (Harvard University Press, 1980); W.G. Lycan (comp.), *Mind and Cognition,* (Blackwell, 1990); David Rosenthal (comp.), *The Nature of Mind* (Oxford University Press, 1991). Putnam ha abandonado ahora el funcionalismo, véase *Representation and Reality* (M.I.T. Press, 1988).

4. Una observación: yo soy uno de los que han sido citados como defensores de esta línea de argumentación. Los otros son Saul Kripke y Frank Jackson.

5. El materialismo eliminativo fue propuesto por primera vez por Paul Feyerabend y Richard Rorty. Más versiones de esta posición son defendidas por Steven Stich, *From Folk Psychology to Cognitive Science: The Case against Belief* (M.I.T. Press, 1983); Paul M. Churchland, *Matter and Consciousness: A Contemporary Introduction to the Philosophy of Mind* (M.I.T. Press, 1984) y Patricia S. Churchland, *Neurophilosophy* (M.I.T. Press, 1986). El comentario de Strawson se encuentra en *Skepticism and Naturalism: Some Varieties* (Columbia University Press, 1985) p. 56.

6. Searle me identifica como un defensor del dualismo de propiedad. Yo prefiero el término "teoría del doble aspecto" para expresar la posición, que proviene de Spinoza, de que los fenómenos mentales son los aspectos subjetivos de estados que pueden ser también descritos físicamente. Pero todo lo que yo sostendría a favor de esta idea es que es algo más aceptable que las otras teorías habituales inaceptables. Comparto la aversión de Searle tanto hacia el dualismo como el materialismo, y no creo que haya ninguna solución al problema mente-cuerpo a la vista.

7. Colin McGinn, en *The Problem of Consciousness* (Blackwell, 1991) sugiere que somos constitucionalmente incapaces de llegar a ese conocimiento. Aunque él pueda estar en lo cierto, creo que su pesimismo es prematuro.

8. Véase el ensayo de Searle, "Minds, Brains, and Programs", en *Behavioral and Brain Sciences* 3, Nº 3 (septiembre de 1980), pp. 417-24. Véase también *Intentionality: An Essay in the Philosophy of Mind* (Cambridge University Press, 1983).

9. Este procedimiento se denomina "algoritmo".

10. La teoría general de este tipo de computabilidad fue desarrollada por Alan Turing.

11. Véase su "Minds, Brains, and Programs".

II

ÉTICA Y FILOSOFÍA POLÍTICA

11

Aristóteles: Sobre *Eudaimonía*[*]

Este ensayo surgió de una réplica a un artículo de John M. Cooper, "Intellectualism and Practical Reasoning in Aristotle's Moral Philosophy", presentado en un encuentro de la Society for Ancient Greek Philosophy, el 28 de diciembre de 1969. Su artículo no fue publicado, pero parte del material apareció más tarde, en su libro Reason and Human Good in Aristotle *(Harvard University Press, 1975).*

La *Ética a Nicómano* muestra una indecisión entre dos enfoques sobre *eudaimonía*: un enfoque comprehensivo y otro intelectualista. Según el enfoque intelectualista, desarrollado en el Libro 10, Capítulo 7, la *eudaimonía* se realiza en la actividad de la parte más divina del hombre, y funciona de acuerdo con su propia excelencia. Se trata de la actividad de contemplación teórica. Según el enfoque comprehensivo (considerado "secundario", en 1178 a 9), la *eudaimonía* implica, esencialmente, no sólo la actividad del intelecto teórico, sino todo el espectro de la vida y acción humanas, de acuerdo con las excelencias más amplias de la virtud moral y la sabiduría práctica. Esta posición conecta la *eudaimonía* con la concepción de la naturaleza humana como compuesta, es decir, involucrando la interacción de razón, emoción, percepción y acción en un cuerpo dotado de alma.

La *Ética a Eudemo* contiene una indecisión similar, expresada en forma menos elaborada. La mayor parte de la obra

* Reimpreso con autorización de *Phronesis 17* (1972).

desarrolla un enfoque comprehensivo, pero el siguiente pasaje aparece en el final:

> Así, esta elección y adquisición de bienes naturales —bienes del cuerpo, riquezas, amigos y otros bienes— que más promueva la contemplación de la divinidad, es la mejor, y esta norma es la más bella; pero aquella que por defecto o por exceso impide vivir y contemplar la divinidad es mala. El hombre posee esto en su alma, y esta es la mejor norma para ella: percibir lo menos posible la otra parte del alma como tal. (1249 b 17-24)*

Aunque este pasaje se encuentra explícitamente en forma algo aislada, parece apoyar la conclusión de que Aristóteles se sentía atraído por un enfoque intelectualista (o quizás espiritualista), sobre los fines de la vida, en la *Etica a Eudemo.* De hecho, considerando el énfasis en el elemento divino en nuestra naturaleza, al final de la *Etica a Nicómano* (1178 b 7-33), no parece fuera de lugar incluir a Dios en la cuestión, hacia el final de la *Etica a Eudemo.* El término "intelectualista" puede ser demasiado seco para expresar los sentimientos casi agustinianos que se encuentran en ambas obras.

Ya que el tema filosófico entre estas dos posiciones surge en virtud de la ambivalencia de la *Etica a Nicómano*, lo analizaré extensamente en este contexto. También comentaré su relación con la psicología del *De anima.* Hay una conexión entre las tendencias intelectualistas en la *Etica*, y la posición de Aristóteles sobre la relación entre el *nous* y el resto del alma. Pero esta última posición me parece que contiene tanta indecisión como la otra. Debido a que él no está seguro de qué es lo que somos, Aristóteles encuentra difícil decir inequívocamente en qué consiste nuestra *eudaimonía* y dónde debe trazarse la línea entre sus constituyentes y sus condiciones necesarias. Más aún, sostendré que el intelectualismo tiene fuertes defensas, aun sin una teoría del ser humano de dos sustancias.

El programa de Aristóteles está expuesto en forma compacta en el Libro 1, Capítulo 7, a partir de 1097 b 22, de la *Etica a Nicómano.* Si no queremos seguir con la trivialidad de que el bien supremo humano es la *eudaimonía*, debemos investigar el *ergon* del hombre, ya que, si algo tiene *ergon*, el bien de ese

* Trad. cast. de EN y EE de Julio Pallí Bonet, Madrid, Gredos, 1985. [T.]

algo es una función de su *ergon*. En general, el *ergon* de algo es lo que hace, gracias a lo cual es lo que es. No todo tiene *ergon*, ya que hay cosas que son, aunque no hagan nada. Pero, cuando algo tiene *ergon*, este determina el bien de esa cosa. El *ergon* propio del hombre, por el cual se mide la excelencia humana, es lo que lo hace ser un hombre, en vez de otra cosa. Los seres humanos hacen muchas cosas, pero, dado que las plantas, peces y animales hacen algunas de estas cosas igualmente bien, o incluso mejor, no se encuentran entre las cosas para hacer en los que consiste ser un humano.

Esto nos pone inmediatamente en dificultades, ya que la inferencia no parece estar fundamentada. Si el rasgo de vida exclusiva de los humanos puede existir en la ausencia de esos rasgos que los humanos comparten con las bestias, el resultado no será un ser humano, sino alguna otra cosa. ¿Y por qué deberíamos considerar el bien supremo de un individuo tan elevado como este, como fin último de individuos complicados y desordenados como nosotros? Uno esperaría, como mínimo, que la *interacción* entre la función que nos diferencia de los animales y las funciones que compartimos con ellos desempeñaría un papel en la definición de *eudaimonía*. Esto implicaría incluir el ejercicio *práctico* de la facultad racional, así como de la contemplativa.

Supongamos que proseguimos con esta crítica, y nos preguntamos si no deberíamos exigir la inclusión de más funciones en la definición del *ergon* humano y, por lo tanto, del bien humano. ¿Qué pasaría, por ejemplo, con la salud o la fertilidad? Se objetaría que estas no pueden cumplir con una condición esencial para ser incluidas: la condición de autonomía.[1] Esta condición establece que los elementos fundamentales del bien humano no pueden deberse meramente a la suerte. Y la salud, por ejemplo, puede ser el resultado solamente de la buena fortuna, más bien que de los esfuerzos de un hombre.

Pero debemos examinar la base de esta objeción. Supuestamente, surge de la condición de que el bien está unido al *ergon* o funcionamiento, y lo que simplemente *le sucede* a una cosa, sea un hombre o una planta, no es una instancia de su funcionamiento correcto o incorrecto. Por lo tanto, si yo contraigo cólera, la intrusión del bacilo agresor es una calamidad, mas no un mal funcionamiento por mi parte, por lo tanto no puede tomarse en cuenta para asignarme *eudaimon* o no. Sin embargo, su *efecto* sobre mí es, ciertamente, un mal funcionamiento. ¿Y por qué este mal funcionamiento, puramente físico, no pue-

de ser *en sí mismo* una desviación de mi bienestar humano (y no sólo debido a sus perjudiciales efectos sobre la vida de la razón)? Si no lo es, sobre la base de que sudar, vomitar y temblar por la fiebre no son cosas que *yo haga*, entonces la pregunta se da por sentada. Dado que la cuestión es, precisamente, si la explicación de lo que la persona es y hace debe incluir o excluir las funciones corporales que comparte con animales e incluso con las plantas. Si digerir, por ejemplo, es algo que hace una almeja, ¿por qué no es algo que un humano también hace es decir, algo que hace, que es parte de lo que es ser un humano, aunque no requiera ningún *esfuerzo*?

Ni en *De anima,* ni en la *Ética a Nicómano* se excluye el elemento nutricional del alma humana, sin embargo, no es uno de los aspectos del funcionamiento humano que Aristóteles esté dispuesto a considerar como medida de *eudaimonía.* Esta posición posee un gran atractivo intuitivo. Si pudiéramos saber por qué la nutrición tiene un rango tan bajo, podríamos tener un indicio de cómo Aristóteles está tentado de dejar de lado todo, excepto el intelecto, hasta lograr que la única cosa intrínsecamente relacionada con la *eudaimonía* sea la cualidad de la actividad contemplativa.[2]

Permítaseme introducir un ejemplo familiar. Un artefacto que es, a la vez, sacacorchos y destapador de botellas tiene la función de sacar corchos y tapas de botellas. Este es un *ergon* simple, que nos permite evaluar su implementación en términos de su capacidad de desempeño exitoso. Sin embargo, no parece lo suficientemente simple como para evitar la pregunta planteada por Aristóteles. Saca las tapas de las botellas, por cierto. Pero, ya que tiene esta función en común con cualquier destapador común, *ese* no puede ser el *ergon* especial (*to idion*) de nuestro artefacto, es decir, el *ergon* por el cual evaluamos su excelencia. De modo que, por descarte, *to idion* tiene que ser sacar corchos. Lamentablemente, esta es una capacidad que comparte con sacacorchos comunes, o sea que tampoco puede ser parte de su *ergon* especial. Obviamente, este argumento no es nada bueno. El artefacto debe tener un *ergon* conjuntivo simple, y su excelencia debe ser una función de ambos miembros de la conjunción.

¿Por qué no puede funcionar una reducción de este tipo para el caso del alma humana y de la *eudaimonía*? Hemos descartado la digestión y la procreación, sobre la base de que las alme-

jas también lo hacen. La sensación y el deseo son comunes con perros. O sea que nos quedamos con la razón. Pero los dioses tienen razón, sin tener las otras facultades, por lo que *ella* tampoco es el *ergon* peculiar del hombre. Por lo tanto, debemos abandonar este método de argumentar por descarte, y reconocer que el hombre tiene un *ergon* conjuntivo, que se superpone con los *erga* de los dioses y perros, así como un sacacorchos y destapador combinado comparte las funciones de sacacorchos y destapadores. Estas se encuentran casualmente en la misma caja de *ergon*.

Si este argumento es correcto, debería apoyar no sólo una posición comprehensiva que abarque las áreas no intelecuales de la conciencia y actividad, sino la inclusión de todas las funciones vitales más bajas en la medida de la excelencia humana. Pero, de hecho, la imagen conjuntiva de las facultades que componen el alma humana es absurda, y si podemos decir por qué es absurda, podremos entender por qué Aristóteles le da a la razón el título de e*rgon idion* del hombre, a pesar de que puede ser, como la digestión, compartido por otros seres.

La idea que opera acá es evidentemente la de una jerarquía de facultades. Las facultades vitales de un organismo complejo no se encuentran todas en el mismo nivel: algunas sirven de soporte para otras. Esto no es tan fácil de explicar claramente. Tomemos, por ejemplo, la subordinación *prima facie* de la nutrición respecto de la percepción y locomoción, que podría ser una razón para rechazar que la nutrición es parte del e*rgon* especial de un animal superior. Pero la percepción y la locomoción de una jirafa, por ejemplo, sirven ampliamente a los fines de la nutrición y la reproducción, de modo que la situación permanece oscura. ¿Qué es lo esencial de ser una jirafa? Una jirafa tiene cierto tipo de vida activa, basada en complejos procesos metabólicos, digestivos y circulatorios, y ordenada de modo de permitir estos procesos para poder funcionar eficientemente. Una cosa está clara: caminar, ver y digerir no son tres actividades separadas que van una al lado de la otra en el mismo individuo, como una muñeca que llora, hace pis y cierra sus ojos. Una jirafa es un organismo, y sus funciones están organizadas en forma coherente. Su excelencia propia no es meramente la conjunción de las excelencias especiales de sus funciones componentes, sino el funcionamiento óptimo del sistema total en la *vida* de la jirafa. Y el nivel más alto para su

explicación no se limitará a la presión sanguínea y peristalsis, sino a la actividad, parte de la cual contribuye en el control de la presión sanguínea y provee material para la peristalsis.

La principal diferencia entre un ser humano y una jirafa es que el ser humano tiene una razón, y que el complejo total de funciones orgánicas tienen tanto actividad racional como irracional. Hay, naturalmente, también retroalimentación: en los humanos, no sólo la percepción y locomoción, sino también la razón, se utilizan para la nutrición y reproducción. La razón también se utiliza en el control de la percepción, locomoción y deseo. No obstante, el enfoque del nivel superior de la vida humana pone todas las otras funciones en una posición de apoyo respecto de la actividad racional. Y aunque la razón nos ayuda a conseguir comida y a movernos, no está sobordinada a esas funciones inferiores. A veces, tiene que servir como conserje o mediador de las pasiones, pero eso no es básicamente *para* lo que sirve. Según una posición admisible, la razón, a pesar de su continuo servicio para con las funciones inferiores, es lo que la vida humana es. Las funciones inferiores le sirven, le dan un contexto y se encuentran, hasta cierto punto, bajo su control, pero la caracterización dominante de un ser humano debe referirse a su razón. Es por eso que Aristóteles se siente atraído por el intelectualismo, y que una posición conjuntiva, que permita incluir varios otros aspectos en la medida del bien, resulta menos aceptable. Ni una posición conjuntiva, ni una posición disyuntiva de la *eudaimonía* es adecuada para estos hechos. El bien supremo del hombre debe medirse en términos de aquello alrededor del cual se organizan todas las otras funciones humanas.

Pero en este punto es indispensable recordar que muchos de los usos prácticos de la razón están al servicio de las funciones inferiores. ¿Se trata de un ejercicio propio de esa facultad, o llega a su límite más allá de los usos de inteligencia, prudencia y coraje, más allá del cálculo racional de la manera más adecuada de gastar nuestro tiempo y dinero o de organizar la sociedad? Esta pregunta lleva a Aristóteles a pasar de la vaga caracterización de que la vida humana, en tanto opuesta a otra vida, es racional, a una consideración de los *objetos* más adecuados para el ejercicio de esa capacidad. Esta es una cuestión peculiar, que no nos planteamos respecto de las jirafas. La actividad no racional de una jirafa y su retroalimentación sobre las funciones metabólicas parece a primera vista un modelo

perfectamente satisfactorio, y podríamos sentirnos tentados de aplicárselo a los seres humanos. ¿Por qué no podemos dar una explicación paralela del funcionamiento humano adecuado, simplemente agregándole razón en la cima de la clase de facultades que cooperan de forma organizada para prolongar la vida del individuo? ¿Por qué deberíamos poner en duda que el uso de la razón para ganarse la vida o para conseguir comida pertenezca a la función central del hombre?

La respuesta es la siguiente: las posibilidades humanas revelan que la razón tiene un uso más allá del ordenamiento de la vida práctica. El círculo de apoyo mutuo entre razón, actividad y nutrición no está completamente cerrado. De hecho, todo él, incluyendo el uso práctico de la razón, sirve para apoyar al individuo en una actividad que trasciende completamente estas cuestiones mundanas. El modelo de retroalimentación no funciona para el *ergon* de humanos, porque el uso mejor y más puro de la razón no tiene nada que ver con la vida diaria. Aristóteles cree, dicho brevemente, que la vida humana no es lo suficientemente importante como para que los humanos gasten su vida en ella. Una persona debería tratar de trascender no sólo sus preocupaciones prácticas individuales, sino también las de la sociedad o humanidad como un todo.

En la *Etica a Nicómano*, Libro 6, Capítulo 7, cuando argumenta que la *sophia* es el tipo de conocimiento más elevado, y el conocimiento uno de los objetos más elevados, Aristóteles dice (1141 a 21-23): "Sería absurdo considerar la ciencia política o la prudencia como lo más excelente si el hombre no es lo más elevado del universo". Las cuestiones teóricas y prácticas deben competir por la atención de la facultad racional, y la capacidad que les permite a los humanos concentrarse en temas más elevados que ellos mismos, al mismo tiempo los malogra respecto de sus preocupaciones inferiores. La imperfección de las aplicaciones de la razón a las cuestiones prácticas consiste en que estas aplicaciones hacen de la vida humana el objeto primario de atención racional, pues el hombre, con su razón, es la única criatura capaz de concentrarse en lo que es más elevado que él mismo, participando en ello hasta cierto punto. Su tiempo es, por así decir, demasiado valioso para gastarlo en algo tan insignificante como la vida humana.

Esto no significa que no haya ninguna distinción entre excelencia y depravación en el dominio práctico. Es ciertamente mejor ejercitar bien la razón, atendiendo a nuestras necesida-

des y el trato con los otros, es decir, tener la virtud moral, que ejercitarla mal. Pero esta es esencialmente una función de guardián de la razón, en la que ella se ocupa de cuestiones —es decir, los detalles sórdidos de la vida de una persona compleja— mucho más inferiores que las que atendería si tuviera más tiempo y no estuviera tan exigida para manejarse.

Creo que es este el punto de vista dominante en los Capítulos 7 y 8 del Libro 10, en el cual el enfoque intelectualista de la *eudaimonía* recibe su apoyo más contundente. Incluso aquí Aristóteles reconoce una posibilidad de duda. No parece estar seguro acerca de si debe describir el resultado como un bien estrictamente humano. Después de argumentar a favor de la vida contemplativa en base a diversas razones, hace en 1177 b 27 la observación de que esta vida sería más elevada que la vida humana. Esta se alcanza, en virtud no sólo de ser un hombre, sino en virtud de algo divino, en lo que los hombres participan. Sin embargo, este elemento divino, que nos da la capacidad de pensar cosas más elevadas que nosotros mismos, es el aspecto más elevado de nuestras almas, y no estamos justificados para renunciar a sus actividades, a fin de concentrarnos en cuestiones más bajas —nuestra propia vida— salvo que estas sean tan acuciantes que nos impidan llevar a cabo la vida contemplativa. Como dice en 1177 b 33, no debemos escuchar a los que nos incitan a pensar que un humano debe tener pensamientos humanos, y un mortal, pensamientos mortales. Más bien, deberíamos cultivar esa porción de nuestra naturaleza que promete trascender el resto. Si alguien insiste en que el resto pertenece a un enfoque completo de la vida humana, entonces se puede decir, en forma algo paradójica, que el bien humano comprehensivo no es todo, y no debe ser el objetivo humano principal. Debemos identificarnos con la parte más elevada de nosotros antes que con el todo. Las otras funciones, que incluyen el uso práctico de la razón misma, le dan apoyo a la forma más elevada de actividad, pero no pertenecen a su propia excelencia como factores componentes primarios. Esto se debe a que los hombres no son, simplemente, la especie animal más compleja, sino que poseen como virtud esencial la capacidad de trascenderse a sí mismos y ser parecidos a los dioses. En virtud de esta aptitud son capaces de *eudaimonía*, mientras que los animales son incapaces de ella, los niños aún no la han alcanzado, y algunos adultos, como los esclavos, no pueden alcanzarla.

Quizá sea una visión insatisfactoria de la naturaleza humana y, por ello, una visión insatisfactoria de qué significa realizarse para un ser humano. Creo, sin embargo, que es una posición muy convincente, que no tiene que depender de un rechazo a la doctrina hilomórfica del alma. Se la puede desafiar de dos maneras: se puede decir simplemente que el *ergon* del hombre es singular, que es una colección de sub-*erga* sin ordenamiento de prioridad y sustento, como los de un cuchillo de muchas hojas. Esto deja sin aclarar cómo establecer la línea que excluye a la buena digestión como componente de *eudaimonía* (y no meramente una contribución a ella). La segunda manera sería conservar la suposición de que el *ergon* del hombre es uno, pero ofreciendo otro enfoque sobre su organización, según la cual el nivel superior de especificación de las facultades humanas no fuera sólo intelectual, sino que incluyera tanto aspectos teóricos como prácticos. Aunque este me parece que podría ser el camino más promisorio, no voy a desarrollarlo aquí.

Notas

1. Cooper señala la importancia de esta condición. Su artículo cita tres enunciados textuales de ella: EE 1215 a 12-19, NE 1099 b 18-25 y, su versión más explícita, en *Política* 1323 b 24-29.

2. Esta conclusión requeriría una explicación, incluso si fuera correcta alguna interpretación de dos sustancias de la Etica. Es decir, aun si la psicología de la *Etica a Nicómaco* incluye un alma compuesta tanto por una forma o realidad primaria del cuerpo, y un intelecto puro distinto de aquella, todavía deberíamos preguntarnos por qué la función propia de la parte hilomórfica no fue incluida en la medición de *eudaimonía*. Y sospecho que cualquier enfoque que tenga sentido sobre esta restricción, a la luz de la teoría de las dos sustancias, podría aplicarse con el mismo éxito para justificar las restricciones correspondientes sobre lo que cuenta a favor de la *eudaimonía* sobre una teoría de la persona puramente hilomórfica, según la cual el *nous*, en vez de ser otra alma, fuera simplemente el tipo más alto de realidad de primer orden en el conjunto, que es distinguible de las otras partes del alma sólo en el pensamiento, y realmente no *distinto*. Que Aristóteles comparte esta posición se ve en una nota entre paréntesis en *NE* Bk 1, cap. 13, cuando está considerando cuál parte del alma posee la excelencia que se debe identificar con la *eudaimonía*. "Nada importa para esta cuestión si estas se distinguen como las partes del cuerpo y todo lo divisible, o si son dos para la razón pero naturalmente inseparables, como en la circunferencia lo convexo y lo cóncavo" (1102 a 30-34).

12

Rawls: Sobre la justicia*

Este ensayo fue originalmente una reseña de Teoría de la justicia, *de John Rawls (Harvard University Press, 1971 [Trad. cast. México, FCE, 1978]). La crítica más notable, de que el argumento de los dos principios de justicia debe presuponer una concepción liberal del bien, y no basarse solamente en una concepción independiente de lo correcto, también fue defendida por otros autores y ha sido objeto de extensos análisis por parte de los defensores y críticos de Rawls, así como por Rawls mismo. (Véase, por ejemplo, su libro reciente,* Political Liberalism *[Columbia University Press, 1993]. Yo mismo he llegado a la posición de que hay una defensa del liberalismo que no depende de una concepción liberal del bienestar humano, y que puede ser aceptada por quienes adoptan otras concepciones del bien. Esta idea es una de las contribuciones más significativas de Rawls, pero es evidentemente un argumento aún en curso. La posición de Rawls, en tanto la personificación del liberalismo filosófico, seguirá siendo blanco de fuertes ataques, ya que la tolerancia política que se encuentra en el corazón del liberalismo se gana la hostilidad de casi todos los que quieren usar el poder coercitivo público al servicio de sus convicciones más profundas respecto de los fines de la vida y que, por lo tanto, rechazan la restricción que propugna el liberalismo.*

Todavía soy escéptico respecto de la fuerza del método de justificación del contrato hipotético, aunque tiene un atractivo innegable como intento de obtener un modelo diferente de imparcialidad que corresponda al sentimiento moral de equidad. Supuestamente, la posición original permite que nos situemos completamente en el lugar de todos los demás, más bien que meramente imaginarnos

* Publicado en *The Philosophical Review* 82 (1973).

que tenemos la misma probabilidad de ser algún otro. Este experimento mental parece alejarnos de las negociaciones interpersonales que nos llevaban al utilitarismo y que fomentaban un método de comparaciones de a pares, basado en un sistema de prioridades entre necesidades humanas, donde la primera prioridad era lo más urgente, independientemente del número de personas involucradas. Este resultado, sea que se derive de la posición original o no, me parece una idea moral muy importante —que es evidentemente una concepción de lo correcto, antes que de lo bueno— y representa una alternativa atractiva a la maximización agregativa como método de resolver conflictos interpersonales de intereses.

En mi reseña había algo importante que no consideré suficientemente: la concepción de la justicia como equidad de Rawls, en tanto una teoría moral de las relaciones *interpersonales; específicamente, una teoría de las formas aceptables de interdependencia entre las vidas y destinos de las personas involucradas en una empresa de cooperación social. Ahora pienso que el nervio moral de la concepción de Rawls de la injusticia es la idea del beneficio o pérdida que resulta de ciertas relaciones inaceptables con los demás. Se trata de una idea esencialmente causal, más que meramente distributiva. Si alguien obtiene una porción más grande o más pequeña de los bienes de la cooperación social, como efecto de tipos equivocados de causas —causas que son moralmente arbitrarias, y que no justifican sus resultados—, entonces esta persona se encuentra en una relación de ventaja o desventaja injusta respecto de sus pares, y tanto él como los demás querrían que las cosas sean diferentes,* por esa misma razón. *Si tenemos un sentido de justicia, según Rawls, no querremos ganar ni perder a partir de cierta clase de relaciones con los demás, y una sociedad bien ordenada debería intentar una forma de vida que evitara tales relaciones. Todavía me parece que esta idea es moralmente más convincente que su justificación por medio de la posición original. Pero* Teoría de la justicia *es un libro gigantesco; sus diferentes líneas de pensamiento tienen diferentes puntos de partida, y no debemos considerarlo como un defecto que no nos lleven todas en la misma dirección.*

Teoría de la justicia es una obra rica, complicada y fundamental. Tiene una serie de argumentos elaborados y presenta muchos temas para el análisis. Esta reseña se centrará en su contribución a las partes más abstractas de la teoría ética.

El libro contiene tres elementos: el primero es una visión del hombre y de la sociedad, tal como deberían ser. El segundo es una concepción de una teoría moral. El tercero es una cons-

trucción que intenta derivar principios que expresan dicha visión, de acuerdo con métodos que reflejan la concepción de la teoría moral. Rawls retoma la tradición contractualista de la filosofía moral y política. El presenta por primera vez su versión del contrato social, una situación de elección hipotética, llamada posición original, en 1958, y aquí la desarrolla en extenso y en detalle. Su objetivo es proponer una manera de tratar los problemas básicos de la elección social, para la cual no existe ningún método reconocido que logre una solución precisa, a través del poder de un problema paralelo de elección individual, especialmente construido, que puede ser resuelto con las intuiciones y procedimientos más confiables de la prudencia racional.

Si esta empresa ha de tener éxito, y si su solución al problema prudencial, que es más claro, se acepta como solución al problema moral, que es más oscuro, entonces la supuesta correspondencia entre ambos problemas debe soportar un gran peso. Los críticos de la teoría tendieron a disentir con Rawls sobre qué principios serían elegidos en la posición original, pero también es necesario examinar los rasgos de la posición que fueron concebidos para apoyar las elecciones más controvertidas, y preguntar por qué los resultados de una decisión tomada bajo estas condiciones altamente específicas y bastante peculiares deberían confirmar la justicia de los principios elegidos. Esta doctrina de correspondencia es fundamental, a la vez que oscura, y no es fácil extraer su defensa del libro. Un tratamiento adecuado del tema tiene un recorrido bastante extenso. Probablemente lo mejor es empezar con la epistemología moral rawlsiana.

Rawls cree que es más útil investigar los fundamentos de la ética donde los resultados sustantivos éticos, sean mayores. Sin embargo, en la Sección 9, expone una posición general que ayuda a explicar su método. La ética, dice, no puede derivarse ni de axiomas autoevidentes ni de definiciones, sino que debe ser desarrollada, como cualquier otro tema científico, a través de la constante interacción entre construcción teórica y observación particular. En este caso, las observaciones particulares no son experimentos, sino juicios morales sustantivos. Es un poco como la lingüística: la ética explora nuestro sentido moral, así como la gramática explora nuestra competencia lingüística.[1]

El intuicionismo trata de captar el sentido moral resumiendo nuestras intuiciones morales particulares en principios de

máxima generalidad, apoyándose en intuiciones para dirimir conflictos entre esos principios. Esto no es lo que Rawls quiere. El intento es que los principios subyacentes posean credibilidad moral intuitiva por sí mismos, y que la teoría total no resuma meramente sino que ilumine y haga creíbles los juicios particulares que ella explica. Más aún, su credibilidad intrínseca debe persuadirnos de modificar o extender nuestras intuiciones, alcanzando así una coherencia teórica mayor. Nuestro conocimiento de hechos contingentes sobre naturaleza y sociedad humana desempeñan un papel sustancial en este proceso.

Cuando esta interacción entre lo general y lo particular produce un resultado relativamente estable, y no parece haber ningún progreso inmediato en ninguno de los dos niveles, se dice que nuestros juicios están en un *equilibrio reflexivo.* Este concepto sugiere que el estado está siempre sujeto a cambiar, y que nuestra mayor aproximación actual a la verdad puede eventualmente ser superada. El artículo indefinido en el título de la obra (*A Theory of Justice)* es significativo: él cree que todas las teorías actuales "son primitivas y tienen graves defectos" (p.72). Sus propios resultados son provisorios. "Dudo" dice Rawls (p. 642) "de que los principios de justicia (tal como los he definido) sean la concepción preferida en relación con algo semejante a una lista completa."

Si los principios y juicios de una teoría son controvertidos y no exigen aceptación intuitiva inmediata, entonces el apoyo que tienen por parte de la concepción moral subyacente cobra especial importancia. Hasta cierto punto, esa concepción puede revelarse directamente en los principios básicos de la teoría, pero es más claramente visible cuando la teoría contiene el modelo o construcción que da cuenta de los principios y de su mutua relación. Teorías alternativas sobre la justicia se representan intuitivamente con diferentes modelos (el utilitarismo, por ejemplo, con el observador imparcial benévolo). El modelo rawlsiano es la posición original, y los principios que utiliza son controvertidos. Para hacerlos más atractivos, la construcción debe expresar una idea intuitiva con credibilidad independiente. Antes de volver al modelo mismo, será útil revisar brevemente las conclusiones sustantivas de la teoría, identificando sus elementos controvertidos y los aspectos que necesitan mayor apoyo independiente.

La doctrina sustantiva rawlsiana es una forma más bien pura de liberalismo igualitario, cuyos elementos controvertidos son su igualitarismo, antiperfeccionismo y antimeritocracia, la primacía de la libertad, y el hecho de que las instituciones sociales se miden, no por su tendencia a maximizar la suma o promedio de ciertas ventajas, sino por su tendencia a compensar las desigualdades naturales que provienen del nacimiento, talento y circunstancias, poniendo estos recursos al servicio del bien común. El bien común se mide en términos de una clase muy limitada de beneficios para los individuos: libertad personal y política, ventajas económicas y sociales, y autorrespeto.

La justicia de las instituciones depende de su conformidad con dos principios. El primero requiere una misma y mayor libertad compatible con una libertad equivalente para todos. El segundo (el principio de diferencia) permite sólo aquellas desigualdades en la distribución de ventajas económicas primarias sociales que beneficien a todos, en particular, al peor situado. La libertad es prioritaria, en el sentido de que no puede ser sacrificada por ventajas económicas y sociales, a menos que estas sean tan escasas o desiguales que impidan el ejercicio significativo de una misma libertad, hasta que las condiciones materiales hayan mejorado.

Esta concepción se opone firmemente a la mera igualdad de oportunidades, la que les permite demasiada influencia a las contingencias moralmente irrelevantes de nacimiento y talento; también se opone a incluir los logros culturales o intelectuales de una sociedad entre los beneficios que harían justo el sacrificio de los bienes más primarios. Lo que importa es que todos tengan las condiciones básicas para la realización de sus propios objetivos, independientemente del nivel absoluto de logros que esto represente.

Cuando sus implicancias sociales y políticas se desarrollan en detalle, como ocurre en la segunda parte del libro, esta posición se vuelve extremadamente atractiva, pero está lejos de ser autoevidente. Al considerar sus bases teóricas, deberíamos preguntarnos si el enfoque contractualista, establecido en términos de la posición original, no depende de suposiciones menos controvertidas que las conclusiones sustantivas que, según se aduce, aquel fundamenta.

La noción de que un contrato es el modelo adecuado para una teoría de la justicia social depende de la posición de que es

justo exigirle a la gente que se someta a procedimientos e instituciones sólo si, dada la oportunidad, hubieran podido estar de acuerdo con antelación sobre los principios a los que se deben someter. Por esta razón, Rawls llama a la teoría "justicia como equidad". (Efectivamente, él cree que una base contractual similar se encuentra entre los principios de moral individual, dando lugar a una teoría de lo correcto como equidad.) La actitud fundamental hacia personas de las que depende la justicia como equidad es el respeto de su autonomía o libertad.[2] Dado que las instituciones sociales existen y la gente nace en ellas, el sometimiento a ellas no puede ser literalmente voluntario, sino que "una sociedad que satisfaga los principios de la justicia como equidad se acerca en lo posible a un esquema voluntario, ya que cumple los principios que consentirían personas libres e iguales bajo condiciones que son imparciales" (p. 30-31).

Antes de considerar si la posición original incluye estas condiciones, debemos preguntarnos por qué se considera que el respeto a la libertad de los demás y el deseo de hacer a la sociedad lo más parecido posible a voluntaria son la fuente principal del sentido de justicia. Esto le da a la libertad una posición de gran importancia desde el comienzo, importancia que conserva en la teoría sustantiva resultante. Pero debemos preguntarnos por qué el respeto hacia la autonomía, por sí mismo, debería dar también mayores resultados.

Cuando se justifica una política sobre la base de que las partes afectadas hubieran (o, incluso, han) estado de acuerdo con ella, gran parte de la justificación depende de las razones para su acuerdo. Si este está motivado por ignorancia, temor, impotencia o un sentido defectuoso de lo razonable, entonces el acuerdo previo, actual o posible, no sanciona nada. En otros casos, podemos obtener o presuponer el acuerdo previo por las razones adecuadas, pero no es el acuerdo lo que justifica lo acordado, sino más bien aquello que justifica el acuerdo mismo. Si, por ejemplo, se debe acordar ciertos principios porque son justos, no puede ser esta la razón que los hace ser justos. En muchos casos, el atractivo del acuerdo previo hipotético es de este carácter. No se trata de una justificación última, ni una señal de respecto hacia la autonomía, sino meramente una manera de recordarle a alguien la clase de juicio *moral* que él adoptaría en la ausencia de influencias distorsivas que surge de esta situación especial.

El consentimiento real o presumible puede ser la *fuente* de una justificación, si y sólo si, se acepta de antemano que las partes afectadas deben ser tratadas como ciertas razones inclinarían a cada una de ellas a querer ser tratadas. Las circunstancias de consentimiento deben poner esas razones en funcionamiento, suprimiendo consideraciones irrelevantes, y el hecho de que se haría esa elección es una razón más para adherir al resultado.

Cuando los intereses de las partes no coinciden naturalmente, se puede igualmente mantener una versión del consentimiento si las partes pueden estar de acuerdo de antemano sobre un procedimiento para resolver conflictos. Pueden estar de acuerdo unánimamente en que el procedimiento las trata equitativamente en aspectos relevantes, aunque no puedan estar de acuerdo de antemano sobre todas las distribuciones de ventajas particulares que puedan resultar. (Un ejemplo de esto sería una lotería que determine quién recibe algún beneficio indivisible.)

Para que el resultado de una elección de este tipo sea moralmente aceptable, dos cosas deben ser verdaderas: (a) la elección debe ser unánime; (b) las circunstancias que posibilitan la unanimidad no deben impedir la igualdad de las partes en otros aspectos. Presumiblemente, deben estar privados de ciertos conocimientos (por ejemplo, quién ganará la lotería) a fin de alcanzar el acuerdo, pero es esencial que no estén privados en forma desigual (como sería el caso, por ejemplo, de someter la disputa a un árbitro, mientras algunas de las partes no saben que es un árbitro extremadamente parcial).

Cuanto más dispares sean los intereses en conflicto, tanto más deben estar las partes privadas de información, a fin de asegurar unanimidad. Entonces surgen dudas respecto de si cualquier procedimiento es confiable para tratar a todos equitativamente con respecto a los intereses relevantes. Hay entonces una cuestión real acerca de si la elección hipotética bajo condiciones de ignorancia, como una representación de consentimiento, puede, por sí misma, dar una justificación moral de resultados que no serían acordados unánimamente si se los conociera con anticipación.

¿Se puede usar un procedimiento de este tipo para justificar principios que evalúen la estructura básica de las instituciones sociales? Evidentemente, las preferencias de los individuos

son tan divergentes, que no estarían de acuerdo voluntariamente sobre una clase común de principios si todos tuvieran la misma voz. Sin embargo, según la teoría de la posición original, se puede apelar al acuerdo previo, exigiendo que la elección hipotética se tome sobre la base de razones que tengan todos los hombres en común, dejando de lado aquellas que los llevarían a elegir diferentes principios e instituciones. No obstante, al restringir la base del acuerdo hipotético de este modo, este puede perder parte de su poder de justificación. Por lo tanto, debemos examinar cuidadosamente las condiciones impuestas a la elección en la posición original. Ya que Rawls no ofrece, en ningún caso, un argumento abstracto a favor del enfoque contractualista, debemos encontrar su defensa en su aplicación.

Se supone que la posición original es la interpretación más favorecida desde el punto de vista filosófico de un *statu quo* inicial hipotético, en el cual ciertos acuerdos fundamentales serían justos. Se puede apelar entonces a los acuerdos en disputas sobre la justicia de las instituciones. Las partes tienen igual voz y eligen libremente; de hecho, ellas pueden llegar independientemente a las mismas conclusiones. Más aún, cada uno de nosotros puede ubicarse en la posición original en cualquier momento, observando sus restricciones bastante especiales sobre argumentos, y eligiendo principios desde ese punto de vista.

Todo esto es posible, porque las bases de la elección tienen severas restricciones, tal como: las partes son mutuamente desinteresadas, es decir, no son ni altruistas ni envidiosas. Sólo saben, de sus propios deseos, lo que es verdad para todos: que tienen algún plan de vida o concepción del bien y un compromiso personal respecto de ciertos individuos. Cualesquiera sean los detalles, todos saben que esos intereses pueden satisfacerse con el uso de bienes primarios muy básicos, bajo condiciones de libertad. También poseen conocimientos generales sobre economía, política y sociología, y también saben que se dan circunstancias de justicia, intereses en conflicto y escasez moderada. Finalmente, creen tener un sentido de justicia que los ayudará a adherir a los principios seleccionados, pero saben suficientemente de psicología moral como para darse cuenta de que sus elecciones deben tener en cuenta las tensiones de compromiso que se sentirían si se adoptaran realmente los prin-

cipios, y la importancia de elegir principios que, al ser aplicados, evoquen su propio apoyo, y con ello tengan estabilidad psicológica. Cualquier otra cosa —su talento, posición social, incluso la naturaleza general o la etapa de desarrollo de su sociedad particular— se encuentra cubierta por un grueso velo de ignorancia, porque es moralmente irrelevante. La elección no puede estar sujeta a influencias de contingencias sociales y naturales que llevarían a algunas partes a hacer presión por ventajas especiales o tener algún poder de regateo especial.

Rawls sostiene que estas restricciones "reúnen, bajo un mismo concepto una cantidad de condiciones sobre los principios que estamos dispuestos, mediante una debida deliberación, a reconocer como razonables" (p. 39), y dice, "se argumentará partiendo de premisas débiles, aunque ampliamente aceptadas, hacia conclusiones más específicas. Cada una de las suposiciones deberá ser, por sí misma, natural y creíble; algunas de ellas pueden incluso parecer inocuas o banales. El objetivo del enfoque contractual es el de establecer que, al considerarlas conjuntamente, imponen límites significativos a los principios aceptables de justicia" (p. 35-36).

No creo que las suposiciones de la posición original sean débiles, inocuas o incontrovertibles. De hecho, la situación, construida de este modo, puede no ser justa. Rawls dice que el objetivo del velo de ignorancia es "eliminar aquellos principios que sería racional proponer para que su aceptación, no obstante la poca probabilidad de éxito que tuvieran, si es que supiéramos ciertas cosas que son irrelevantes desde el punto de vista de la justicia" (p. 36). Aceptemos que las partes deberían ser iguales y no poseer información que las lleve a buscar ventajas sobre la base de factores moralmente irrelevantes como raza, sexo, origen o dotes naturales. Pero tampoco poseen información sobre su concepción particular del bien. Parece extraño considerar a esta última como moralmente irrelevante desde el punto de vista de la justicia. Si alguien elige ciertos principios, sobre la base de su concepción del bien, no estará buscando ventajas especiales para él, dado que no conoce su lugar en la sociedad. Más bien, estará optando por principios que favorecen el bien para todos, según como su concepción lo define. (Supongo que una concepción del bien es solamente eso, y no simplemente un sistema de gustos o preferencias.) Sin embargo, Rawls parece creer que sería injusto permitirles a las personas que

presionaran por la realización de su concepción del bien, así como permitirles que lo hicieran por la ventaja de su clase social.

Es verdad que las diferentes concepciones del bien dividen a los hombres y producen conflictos, de modo que permitirles poseer este conocimiento en la posición original impediría la unanimidad. Rawls concluye que se debe suprimir esta información, y sustituirla con una idea común, que les permita el consenso sin elegir ninguna concepción particular del bien. Esto se logra por medio de la clase de bienes primarios que se supone racional desear, independientemente de lo que uno desee. Otra conclusión posible, sin embargo, es que el modelo de la posición original no funciona, porque, a fin de asegurar unanimidad espontánea y evitar la necesidad del regateo, se debe suprimir información moralmente relevante y, más aún, suprimirla de una manera que no trata equitativamente a las partes.

Lo que Rawls quiere hacer, al usar la noción de bienes primarios, es dar un punto de Arquímedes, como él lo llama, desde el cual la elección sea posible sin injusticia respecto de las concepciones más completas del bien que llevan a la gente a diferir entre sí. Se presupone una *teoría* del bien, pero que es ostensiblemente neutral entre diferentes concepciones particulares, y otorga por lo menos un denominador común, sobre el cual se puede basar una elección en la posición original, sin injusticia respecto de alguna de las partes. Sólo más tarde, una vez que se alcanzaron los principios de justicia sobre esta base, será posible dejar de lado ciertos intereses o metas particulares, en tanto ilegítimos, porque son injustos. Es un rasgo fundamental de la concepción de la equidad de la posición original, no permitir la elección de principios de justicia que dependan de una concepción particular del bien, sobre la que las partes puedan diferir.

No creo que esta construcción logre tal objetivo, y hay razones para suponer que ella no puede llevarse a cabo exitosamente. Cualquier situación hipotética que requiera un acuerdo entre las partes tendrá que imponer severas restricciones en la elección, y estas restricciones sólo pueden ser justificadas en términos de una concepción del bien. Es uno de esos casos en los cuales no se puede tener neutralidad, porque la neutralidad necesita tanta justificación como cualquier otra posición.

La concepción mínima del bien de Rawls no es un supuesto débil: depende de un fuerte supuesto de que esa concepción reducida es suficiente a los efectos de la justicia. El rechazo a jerarquizar las concepciones particulares del bien implica una muy fuerte tolerancia respecto de las inclinaciones individuales. Rawls se opone no sólo a las concepciones teleológicas, según las cuales la justicia requiere adhesión a los principios que maximicen el bien. También se opone a la posición natural, de que incluso en una teoría no teleológica lo que es justo depende de lo que es bueno, al menos en el sentido de que una concepción correcta del bien tiene que usarse para determinar qué se considera una ventaja y qué una desventaja, y en qué medida, a los propósitos de distribución y compensación. Interpreto que él dice que los principios de justicia son objetivos y reconocibles interpersonalmente, en una forma en que las concepciones del bien no lo son. El rechazo a jerarquizar concepciones individuales y la confianza en los bienes primarios responden al propósito de asegurar esta objetividad.

Es posible que no pueda lograrse objetividad tan fácilmente.[3] La supresión de conocimiento necesaria para lograr unanimidad no es igualmente equitativa para todas las partes, porque los bienes primarios no son igualmente valiosos para la prosecución de todas las concepciones del bien. Ellos servirán para promover muchos planes de vida individuales (algunos en forma más eficiente que otros), pero no servirán para implementar concepciones que sostienen que una buena vida puede ser llevada a cabo sólo en ciertos tipos de estructura social bien definidos, o sólo en una sociedad que funcionan organizadamente para la realización de ciertas facultades humanas superiores y la supresión de las más bajas, o sólo ciertos tipos de relaciones económicas entre personas. El modelo tiene una tendencia fuertemente individualista que está, además, reforzada por los supuestos motivacionales del desinterés mutuo y la ausencia de envidia. Estos supuestos tienen el efecto de excluir las pretensiones de las concepciones del bien que dependen fuertemente de la relación entre la propia posición y la de los demás (aunque Rawls está preparado para permitir tales consideraciones, siempre que no afecten la autoestima). La posición original no parece presuponer una teoría neutral del bien, sino una concepción liberal, individualista, según la cual lo mejor que podemos desearle a alguien es

la prosecución sin obstáculos de su propio camino, siempre que no interfiera con los derechos de los demás. La posición está desarrollada en forma convincente en las últimas partes del libro, pero sin un sentido de su carácter controvertido.

Entre los diferentes planes de vida de este tipo general, la construcción es neutral. Pero, dado que muchas concepciones del bien no calzan en este patrón individualista, ¿cómo podemos describirlo como una situación justa de elección de principios de justicia? ¿Por qué las partes en la posición original deberían estar preparadas a comprometerse con principios que pueden frustrar o contravenir sus convicciones más profundas, sólo porque están privados del conocimiento de esas convicciones?

No parece haber ninguna manera de rediseñar la posición original que suprima una suposición restrictiva de esta clase. Podríamos pensar que es un mejoramiento permitirles a las partes tener información completa sobre las preferencias y concepciones del bien de todas ellas, privándolos solamente del conocimiento de quiénes son. Pero esto, como señala Rawls (pp. 202-204), no daría ningún resultado. Dado que cualquiera de las partes conservaría su concepción del bien, y al elegir desde diferentes puntos de vista, no alcanzarían unanimidad o de lo contrario no tendrían ningún objetivo propio, y deberían tener que elegir en términos de los objetivos de todas las personas que ellas podrían ser: una exigencia ininteligible, que no da ninguna base para una elección unificada, en ausencia de una concepción dominante. Por lo tanto, la reducción a una base común de elección es esencial para que el modelo pueda operar, y la selección de esa base representa, inevitablemente, una suposición fuerte.

Volvamos ahora al argumento que lleva a la elección de los dos principios en la posición original. El núcleo del argumento aparece en las Secciones 26-29, intercalado con un argumento contra la elección del principio de utilidad promedio. Rawls defiende extensamente su discutible afirmación de que, en la posición original, es racional adoptar la regla maximin, que lleva a elegir aquellos principios que favorecen la base de la jerarquía social, en vez de aceptar un riesgo mayor en la base, a cambio de la posibilidad de mayores beneficios en la cima (que sería prudencialmente racional si uno tuviera la misma posibilidad de ocupar cualquier lugar en la sociedad).

Rawls dice (pp. 182-183) que se dan tres condiciones en la posición original, que convierten al maximin en altamente probable. (1) "Tiene que haber alguna razón para no tomar en cuenta la estimación de estas... probabilidades." (2) "La persona que escoge tiene una concepción del bien tal, que le importa muy poco o nada lo que puede ganar por encima del mínimo que seguramente obtiene si sigue la regla maximin." (3) "Las alternativas rechazadas tienen resultados que difícilmente pueden aceptarse." Consideremos estas condiciones a su turno.

La primera de ellas es muy importante, y su tesis de que se cumple en la posición original no se basa simplemente en un rechazo general al principio de razón insuficiente, es decir, el principio según el cual, donde se desconocen las probabilidades, estas serán consideradas iguales. Ya que uno podría caracterizar a la posición original de modo tal que sería prudencialmente racional que las partes eligieran como si tuvieran la misma posibilidad de ocupar cualquier lugar en la sociedad, y el problema consiste en ver por qué esta sería una representación inadecuada de los fundamentos de una elección de principios.

Un factor mencionado por Rawls es que el tema objeto de la elección es extremadamente serio, ya que comprende la elección de las instituciones que van a determinar todo el futuro de la vida de las partes y de sus allegados. No es meramente una elección de alternativas para una ocasión determinada. Ahora, esto sería una razón para una elección conservadora, aun si uno conociera las probabilidades relativas de diferentes resultados. Sería irresponsable aceptar un riesgo ínfimo de un futuro terrible para uno y para sus descendientes, a cambio de una buena probabilidad de riqueza o poder. Pero hay que explicar por qué deberían excluirse totalmente las probabilidades, y no sólo con respecto a los resultados más inaceptables. El principio de diferencia, por ejemplo, se supone que se aplica a todos los niveles de desarrollo social, de modo que no se justifica solamente por el deseo de evitar graves riesgos. El hecho de que están involucradas las perspectivas totales de vida no parece ser una explicación suficiente. Debe haber alguna razón en contra de permitir probabilidades (por ejemplo, probabilidad proporcional al número de personas en cada posición social) en la elección de distribuciones por sobre un mínimo aceptable. Permítaseme remarcar que no estoy planteando una

pregunta sobre la teoría de la decisión, sino sobre el diseño de la posición original y el alcance del velo de ignorancia. ¿Por qué se considera que sólo se alcanza una solución justa si se suprimen estas consideraciones?

Esta supresión está justificada, creo, sólo bajo la suposición de que se consideran moralmente irrelevantes las proporciones de personas en las diferentes posiciones sociales, y esto tiene que ser así porque no se considera aceptable sumar ventajas y desventajas sobre personas, de modo que la pérdida para algunos se compense con la ganancia para otros. Este aspecto del diseño de la posición original parece motivado por el deseo de evitar extenderle a la sociedad como un todo el principio de elección racional para un solo hombre. Ahora bien, se supone que esta es una de las *conclusiones* del enfoque contractual, y no uno de sus supuestos. Sin embargo, las restricciones a la elección en la versión rawlsiana de la posición original están diseñadas para descartar la posibilidad de esa extensión,[4] exigiendo que se dejen de lado las probabilidades. No veo ninguna manera de evitar presuponer alguna concepción definitiva sobre este tema en el diseño de una situación contractual. Si esto es verdad, entonces un enfoque contractualista no puede darle demasiado apoyo a ninguna concepción.

Consideremos la segunda condición. Teniendo en mente que las partes en la posición original no conocen la etapa de desarrollo de su sociedad, y por lo tanto no saben cuál va a ser el mínimo garantizado por una estrategia maximin, es difícil entender cómo un individuo puede saber que "le importa muy poco o nada, lo que puede ganar por encima del mínimo". La explicación que da Rawls (p. 184) parece muy débil. Aun si las partes en la posición original aceptan la prioridad de la libertad, y aun si el velo de ignorancia las deja con una concepción esquelética del bien, parece imposible que les importe muy poco el aumento de los bienes primarios económicos y sociales por sobre lo que garantiza el principio de diferencia en cualquier etapa del desarrollo social.

Finalmente, la tercera condición, que deben dejarse de lado ciertas posibilidades por ser inaceptables, es ciertamente una razón para exigir un mínimo social y la prioridad de las libertades personales básicas, pero no es una razón para adoptar la regla maximin en esa forma general necesaria para justificar la elección del principio de diferencia. Esta debería apoyarse en premisas igualitarias más fuertes.[5]

Algunas de estas premisas aparecen en otras partes del argumento. Por ejemplo, la idea fuertemente igualitaria de que el sacrificio en la base es siempre peor que el sacrificio en la cima desempeña un papel central en la apelación a tensiones de compromiso y estabilidad psicológica. Rawls la utiliza contra la alternativa utilitarista al principio de diferencia, por ejemplo, que los sacrificios que podría exigir el utilitarismo serían psicológicamente inaceptables.

> Los principios de la justicia se aplican a la estructura básica del sistema social y a la determinación de las perspectivas de vida. Precisamente lo que impide el principio de utilidad es un sacrificio de estas perspectivas. Aunque seamos los menos afortunados, habremos de aceptar que las mayores ventajas de los demás son razón suficiente para tener menores expectativas en el curso de nuestra vida. Es, sin duda, una exigencia extrema. De hecho, si se concibe a la sociedad como a un sistema de cooperación para promover el bien de sus miembros, parece casi increíble el que se espere que algunos ciudadanos hayan de aceptar, con base en principios políticos, tener todavía menores perspectivas de vida en aras de los demás. (p. 207)

Nótese que si sustituimos las palabras "principio de diferencia" por "principio de utilidad", tenemos un argumento que alguien podría usar en contra del principio de diferencia, apelando a los sacrificios que este exige que hagan aquellos situados en la cima del orden social. Ellos tienen que vivir en instituciones que limiten sus perspectivas de vida, salvo que una ventaja para ellos beneficie también a los peor situados. La única diferencia entre ambos argumentos se encuentra en la posición relativa de las partes y sus sacrificios.[6] Naturalmente, se trata de una diferencia esencial, pero depende de un juicio moral, es decir, que los sacrificios que disminuyen la desigualdad social son aceptables, en tanto que los sacrificios que favorecen la desigualdad no lo son.

Por lo tanto, esta apelación a la estabilidad psicológica y a las tensiones de compromiso agrega a las razones de la selección en la posición original una concepción moral que pertenece a la teoría sustantiva. Es posible que el argumento reciba algún apoyo por medio de la idea rawlsiana del desarrollo natural de los sentimientos morales, pero estos, a su vez, no son independientes de su teoría ética. Si una elección hipotética en

la posición original tiene que basarse en lo que pretendemos considerar moralmente aceptable en la vida real, entonces no es esa elección la base verdadera de su aceptación.[7]

Otra conclusión fuerte de la teoría es la prioridad de una misma libertad, expresada en el ordenamiento lexical de ambos principios. El argumento a favor de una libertad *igual* como meta natural es directo. No se puede aplicar permanentemente ningún análogo del principio de diferencia a la libertad, porque esta no puede aumentarse indefinidamente. Se llega a un punto donde el aumento de la libertad del peor situado puede lograrse, no mediante el aumento de la libertad del mejor situado, sino eliminando la brecha entre ambos. Si se trata de maximizar para todos lo que realmente tiene un máximo, el resultado es la igualdad.

Sin embargo, la prioridad de la libertad sobre otros bienes se elige en la posición original sobre la base de un juicio de que el interés fundamental en la determinación del propio plan de vida se vuelve prioritario una vez que las necesidades más básicas han sido satisfechas, y que el valor de más aumentos en otros bienes depende primariamente de la habilidad de utilizarlos bajo las condiciones de máxima libertad. "Por lo tanto, el deseo de libertad es el interés regulatorio principal que las partes deben suponer que todas tienen en común. El velo de ignorancia las obliga a abstraerse de los cursos particulares de sus planes de vida, llevándolos así a esta conclusión. Se deriva de ello el orden serial de los dos principios" (p. 543 de la ed. en inglés). Las partes consideran además que una misma libertad garantiza a todos una autoestima básica, con lo cual serán aceptables algunas diferencias en posición social y salud. Nuevamente, aquí hay una concepción del bien individual explícitamente liberal, que se utiliza para defender una elección en la posición original.

He intentado argumentar que el método contractual que utiliza Rawls tiene presupuestos suficientemente sólidos y que, consecuentemente, la posición original apoya sus conclusiones en forma menos independiente de lo que parece al principio. El liberalismo igualitario que Rawls desarrolla y la concepción del bien de la que depende son muy convincentes, pero la posición original sirve para modelarlos, más que para justificarlos. El enfoque contractualista, junto con una concepción no liberal del bien, conllevaría resultados diferentes, y algunas concep-

ciones del bien son incompatibles con cualquier enfoque contractualista de la justicia. Creo que las conclusiones de Rawls pueden defenderse convincentemente con argumentos morales directos a favor de la libertad e igualdad, algunos de los cuales Rawls desarrolla, y otros se encuentran representados indirectamente en su enfoque actual sobre las bases y condiciones de elección en la posición original. El destaca que

> ...es digno de hacerse notar desde el comienzo que la justicia como equidad, igual que otros puntos de vista contractuales, consiste en dos partes: (1) una interpretación de la situación inicial y del problema de elección que se plantea en ella, y (2) un conjunto de principios en los cuales, se dice, habrá cierto acuerdo. Se puede aceptar entonces la primera parte de la teoría (o una variante de ella) sin aceptar la otra, y viceversa. (p. 33)

Rawls sugiere que es más probable rechazar los principios que su base contractual, pero yo creo que es al revés. Me parece probable que, en el largo plazo, este libro alcance un lugar permanente en la bibliografía de teoría política, debido a la doctrina sustantiva que desarrolla en forma tan elocuente y convincente. La credibilidad de sus resultados no deberá ser tomada para confirmar la validez del método, aunque tales inferencias no siempre son correctas. Tal vez se pueda alcanzar la solución a los problemas combinatorios de la elección social por medio de una elección autointeresada bajo condiciones de incertidumbre cuidadosamente especificadas, pero la base para una solución de este tipo todavía no ha sido descubierta.

Es un libro ya famoso e influyente, y es inevitable que ocupe por bastante tiempo a estudiantes de filosofía, política, derecho y economía. Cuán larga será la vida de una obra y cuán fuerte será su impacto sobre los hábitos intelectuales de personas reflexivas es una pregunta que nunca puede responderse con exactitud, pero es una pregunta interesante. Aunque *Teoría de la justicia* es, en su mayor parte, fácil de leer, no posee la distinción literaria que ayudó a incluir a otras obras políticas importantes —las de Hobbes o Mill, por ejemplo— en la propiedad intelectual común de la humanidad. Sin embargo, sí posee otro rasgo muy importante. Leerlo es una experiencia

muy fuerte, porque uno se encuentra en contacto directo, en cada punto, con un temperamento y una mentalidad sorprendentes. Es, en ese sentido, una obra muy personal, y las percepciones y actitudes que se encuentran en ella son vívidas, inteligentes y atractivas. La perspectiva del libro no es característica de su tiempo, porque no es pesimista, alienado, enojoso, sentimental ni utópico. En cambio, conlleva algo que hoy puede parecer increíble: una esperanzada afirmación de las posibilidades humanas. Sin embargo, la esperanza tiene un fundamento, porque Rawls posee un profundo sentido de las múltiples conexiones entre las instituciones sociales y la psicología individual. Sin engañarse, describe un orden social pluralista que convocará el apoyo de los hombres libres y evocará lo mejor de ellos. Haber hecho de esta visión algo preciso, vivo y convincente es un logro memorable.

Notas

1. Esta analogía me parece falsa, porque las intuiciones de los hablantes de lengua materna no son decisivas en cuanto a la gramática. El acuerdo entre los hablantes de lengua materna, cualquiera sea, es inglés, pero sea cual fuere el acuerdo entre las personas en condenar un acto, no significa que sea necesariamente equivocado. Por lo tanto, la posibilidad intrínseca de una teoría ética puede producir un cambio en nuestras intuiciones morales. Nada de esto ocurre en la lingüística (véase la sugerencia de Rawls, en p. 69), donde el test final de una teoría es su habilidad para explicar los datos.

2. Rawls, extendiéndose sobre este punto, dice que su posición puede recibir una interpretación kantiana, pero los detalles de esta analogía no son siempre convincentes. Véase, por ejemplo, en página 288, su afirmación de que los principios de justicia son imperativos categóricos, porque los argumentos a favor de ellos no suponen que las partes intervinientes en el acuerdo tengan fines particulares, sino sólo que desean aquellos bienes primarios que es racional querer, independientemente de lo que uno desee. Primero de todo, el deseo de esos bienes primarios no es, en sí mismo, el motivo para obedecer los principios de justicia en la vida real, sino sólo para elegirlos en la posición original. Segundo, los imperativos que se derivan de tal deseo deberían ser hipotéticos y asertivos en el sistema de Kant, y no categóricos. Pero, dado que se supone que nuestra adhesión a los dos principios está motivada por un sentido de justicia mayor a nuestra gratitud respecto de los beneficios que recibimos de instituciones justas, los imperativos de justicia como equidad parecerían ser, de hecho, hipotéticos y problemáticos (*Fundamentación de la metafísica de las costumbres,* pp. 415-46 de la Prussian Academy Edition).

3. Le debo a Mary Gibson las ideas de este párrafo.

4. Esto es, no rechazan la suposición de que la extensión es aceptable: suponen que es inaceptable.

5. Un factor no considerado en el argumento de Rawls, que sugiere que el principio de diferencia puede ser demasiado débil, es el siguiente: si se permiten beneficios diferenciales sociales y económicos para proporcionar incentivos, entonces los que están en la cima tenderán a ser aquellos que tienen ciertos talentos y habilidades, y los que están en la base, aunque estén mejor de lo que estarían de otro modo, no tendrán, en general, esas cualidades. Un programa tan coherente de premios inevitablemene afecta el sentido del valor intrínseco de las personas, y cualquier sociedad que opere desde el principio de diferencia tendrá un sabor meritocrático. Esto es muy diferente del caso en que una distribución desigual, que beneficia al peor situado, no está correlacionada visiblemente con alguna cualidad independiente. Rawls sugiere (p. 546) que la "envidia justificada" puede explicarse en el funcionamiento del principio de diferencia, mediante la inclusión de la autoestima entre los bienes primarios. Pero no enfatiza las *bases* de la desigualdad de ingresos. El fenómeno que describo no es *envidia.* Rawls está más que dispuesto a confiar en la igual libertad como soporte de la autoestima; esto lo lleva a subestimar el efecto de las recompensas diferenciales sobre el concepto que tiene la gente sobre sí misma. No se percibe meramente como buena suerte a una recompensa que aparece ligada a cierta cualidad.

6. Después de todo, el mismo sacrificio podría ocurrir tanto en la cima como en la base, según la etapa de evolución de la sociedad.

7. Se le puede hacer una objeción similar a la afirmación de Rawls de que el principio de diferencia tiene una condición de ventaja recíproca que permite que todos cooperen voluntariamente en el orden social. Obviamente, los que están abajo no preferirían otro acuerdo, pero, ¿qué ocurre con los que están arriba? Rawls dice lo siguiente: "Así, cuando los más aventajados vean esta cuestión desde una perspectiva general, reconocerán que el bienestar de todos depende de un esquema de cooperación social, sin el cual nadie podría tener una vida satisfactoria; reconocerán también que sólo si los términos del esquema son razonables, podrán esperar la cooperación voluntaria de todos. Por eso se consideran a sí mismos ya compensados, en cierta medida, por ventajas a las cuales nadie (ni siquiera ellos) tenía previamente derecho. Renuncian a la idea de maximizar un valor medio equilibrado y consideran el principio de la diferencia como una base para regular la estructura básica" (p. 125-126). Pero si algún esquema de cooperación social es necesario para que *cualquiera* tenga una vida satisfactoria, todos se beneficiarán con una gran variedad de esquemas. Suponer que el peor situado necesita más beneficios para cooperar voluntariamente, en tanto que el mejor situado no los necesita, es simplemente repetir el principio igualitario.

13

Nozick: Libertarismo sin fundamentos*

Este ensayo fue originalmente una reseña de Anarquía, Estado y utopía, *de Robert Nozick (Basic Books, 1974 [Trad. esp. México, FCE, 1988]). Mi ataque, bastante implacable, era un intento de rechazar el apoyo que, supuestamente, Nozick le ha dado brillantemente a una posición política que todavía me parece sumamente objetable. Desde entonces, Nozick ha modificado su posición (véanse* The Examined Life *[Simon and Schuster, 1989], pp. 286-92; y* The Nature of Rationality *[Princeton University Press, 1993], p. 32), pero su libro sigue siendo la enunciación canónica del libertarismo filosófico, y mientras que el libertarismo radical no es, en sí mismo, un peligro político, algunas de sus ideas morales son un elemento importante del antiigualitarismo moderno del ala derecha. La obra de Nozick revela en todo momento una gran inteligencia y encanto, pero en este caso, lo que era intuitivamente probable para él —algunas de sus premisas no discutidas sobre las que se basa toda la estructura así como sus consecuencias específicas— a mí me parecía totalmente inaceptable.*

No creo que mi reacción haya sido inusitada, pero, aunque el libro haya persuadido sólo a unos pocos, se ha convertido en un clásico, lo que es en sí mismo un hecho notable en filosofía política.

El liberalismo es la conjunción de dos ideales. El primero es el de la libertad individual: libertad de pensamiento, expresión, religión y acción política; libertad de interferencia estatal en la vida privada, personal, y el ejercicio de las inclinaciones

* Reimpreso con autorización de *The Yale Law Journal* 85 (1975).

individuales. El segundo ideal es el de una sociedad democrática, controlada por sus ciudadanos y que sirve a sus necesidades, en la cual las desigualdades del poder político y económico y de la posición social no son excesivas. Las medidas para promover el segundo ideal incluyen impuestos progresivos, garantía pública de un mínimo social, y el aislamiento de los asuntos políticos de la excesiva influencia de la riqueza privada. Acercarse a uno de estos dos ideales es muy difícil. Tratar de alcanzarlos a ambos trae consigo serios dilemas. En tales casos, el liberalismo tiende a darle prioridad al respeto a ciertos derechos personales, aun a costa sustancial en la realización de otros bienes, como la eficiencia, igualdad y estabilidad social.

El más formidable desafío al liberalismo, tanto intelectual como político, proviene de la izquierda, desde donde se argumenta que las grandes salvaguardas de la libertad individual representan un impedimento demasiado grande para el logro de la igualdad económica y social, de un rápido progreso económico para salir del subdesarrollo, y de la estabilidad política. La gran mayoría de las personas en el mundo está gobernada por esta concepción. Tal vez la cuestión más difícil se plantea por el poder económico y la desigualdad política que esto puede originar. La crítica de la izquierda es que no se puede atacar la concentración dañina del poder económico —o es inevitable— salvo que se restrinjan las acciones individuales más fuertemente de lo que el ideal liberal de la libertad personal permite. En una democracia liberal, en la que la riqueza privada controla el proceso político, una redistribución radical es improbable. En contra de esta crítica, se debería desafiar sus afirmaciones fácticas o argumentar que la importancia de la libertad compensa esas desventajas.

El liberalismo también recibe críticas provenientes de la derecha. Los ataques más fuertes no son teóricos: a la derecha, en sus manifestaciones políticas más importantes, no le importa demasiado la libertad individual cuando esa libertad amenaza la distribución desigual de riqueza y poder. Pero también hay un desafío teórico de la derecha, llamado libertarismo, que, aunque no presenta un problema moral serio para los liberales —como sí ocurre con los ataques de la izquierda—, ambos son, de algún modo, simétricos. El libertarismo, como el izquierdismo, se aferra a uno de los dos elementos del ideal liberal, y pregunta por qué su realización debe inhibirse para

satisfacer las demandas del otro. En lugar de adoptar el ideal de igualdad y bienestar general, el libertarismo exalta la pretensión de libertad individual de acción, y se pregunta por qué el poder del Estado ha de ser permitido, aun cuando esa interferencia sea la implantación de impuestos progresivos y la atención pública de salud, educación, y un nivel de vida mínimo.

En *Anarquía, Estado y utopía*, Robert Nozick trata de extender la posición libertaria de modo que convenza a algunos de los que todavía no la adoptaron. A pesar de su ingenuidad de detalle, el esfuerzo es totalmente vano como intento de persuadir, y mucho menos exitoso de lo que podría ser, como intento de explicarle a alguien que no adopta la posición, por qué otra persona sí la acepta. El libro podría ocupar el lugar de un texto oficial de la teoría política libertaria, pero es improbable que logre incorporar nuevos adeptos a dicha posición, a no ser que convierta a algunos pocos anarquistas filosóficos desprevenidos, al convencerlos de que un Estado mínimo no necesita, después de todo, violar sus requisitos morales austeros.

Para presentar un desafío serio a otras posiciones, un análisis sobre el libertarismo tendría que explorar los fundamentos de los derechos individuales, y las razones a favor y en contra de diferentes concepciones acerca de la relación entre esos derechos y otros valores que el Estado puede promover. Pero el libro de Nozick es teóricamente insustancial: no toma los problemas principales; y, por lo tanto, falla al hacer la clase de contribución a la teoría política que se esperaría de alguien con sus logros filosóficos.[1] En el prefacio, Nozick anuncia que se convirtió al libertarismo por la fuerza decisiva de los argumentos (p. ix), pero no hay en el libro tales argumentos. El deja el establecimiento de la fundamentación moral para otro momento, y su breve indicación de cómo defender las posiciones básicas no es satisfactoria. Explicaré a continuación por qué no es probable que su teoría llegue a desarrollos posteriores.

Nozick parte de la premisa indiscutible de que los individuos tienen ciertos derechos inviolables que no pueden ser transgredidos intencionalmente por otros individuos, o por el Estado, para ningún propósito. Se trata de los derechos a no ser matado o atacado si uno no está haciendo ningún daño, a no ser forzado o encarcelado, a que no se robe o destruya la propiedad, a no ser limitado en el uso de la propiedad, siempre que uno no viole los derechos de los demás. Concluye que el

único Estado moralmente permisible sería el Estado mínimo guardián, un Estado limitado a proteger a la gente de asesinatos, ataques, robos, fraude y violación de contratos. El argumento no es de los que obtienen una conclusión sorprendente a partir de premisas admisibles. Nadie (salvo, quizás, un anarquista) que no acepte de antemano la conclusión, aceptaría la premisa, y la inaceptabilidad de una sólo sirve para reforzar la convicción de inaceptabilidad de la otra.

Naturalmente, cualquier oposición al poder de los gobiernos encontrará cierta simpatía por parte de observadores de la escena contemporánea, y Nozick enfatiza la conexión entre su posición y la lucha en contra de la regulación jurídica de la conducta sexual, consumo de drogas y estilos de vida individuales. Es fácil desarrollar una aversión al poder del Estado viendo cómo los Estados lo ejercen actualmente. Sus actividades a menudo incluyen el asesinato, la tortura, la prisión política, la censura, la conscripción para guerra agresiva, el derrocamiento de gobiernos de otros países, sin mencionar las escuchas ilegales, la violación de correspondencia y la regulación de la conducta sexual de ciudadanos.

La objeción a estos abusos, sin embargo, no es que exista el poder del Estado, sino que se lo use para el mal y no para el bien. La oposición a estos males no puede traducirse en una objeción al bienestar, educación pública o impuestos graduales a las ganancias. Un argumento práctico razonablemente convincente para reducir el poder de los gobiernos podría basarse en los resultados negativos de ese poder. Pero es dudoso que un gobierno limitado a las funciones de policía, tribunales, prisiones y defensa nacional fuera muy benigno, o especialmente protector de los derechos individuales.[2] En la práctica, incluiría probablemente la peor parte de lo que tenemos ahora, dejando afuera mucho de lo mejor. Por eso quienes se preocupan por la libertad individual normalmente no se oponen a un gobierno fuerte con poder para promover bienes deseables, siempre que el ejercicio de ese poder esté limitado por fuertes garantías. Los gobiernos deberían promover el bien y prevenir el mal, así como proteger derechos. ¿Cómo estar en desacuerdo con ello?

Si hay una respuesta a esta pregunta, debe provenir de la fundamentación ética de la teoría política. Nozick dice:

> La filosofía moral establece el trasfondo y los límites de la filosofía política. Lo que las personas pueden o no hacerse unas a otras

> limita lo que pueden hacer por medio del aparato estatal, o lo que pueden hacer para establecer dicho aparato. Las prohibiciones morales que es permisible imponer son la fuente de toda legitimidad que el poder coercitivo fundamental del Estado tenga. (p. 19)

Creo que este principio es correcto e importante. El ejercicio del poder del Estado no es la acción de una entidad separada con más derechos morales que los de los individuos, derechos a usar la fuerza contra personas por razones que no justificarían el uso de la fuerza por individuos o grupos de individuos *per se*. Si los gobiernos tienen el derecho a la coerción, debe ser un derecho de las personas, que establecen y apoyan instituciones gubernamentales, y de quienes actúan en su nombre.

Hay un problema en enunciar esta posición de modo tal de evitar la banalidad. Para alguien que cree que los gobiernos tienen derechos mucho más amplios que los individuos, siempre se podría agregar que la existencia de tales derechos implica el derecho individual correspondiente de combinarse con otros para instaurar un gobierno y actuar por medio de él a fin de ejercer derechos más amplios de coerción y control. Pero, según esa posición, estos derechos individuales se derivarían de los derechos del Estado, y no al revés. La posición de Nozick, que parece correcta, es que los derechos y deberes individuales son la base de lo que los gobiernos pueden y deben hacer.

Pero él parece inferir de este principio ético una fuerte consecuencia epistemológica que no posee: que es posible determinar lo que los gobiernos pueden y deben hacer preguntando, en primer lugar, qué pueden hacer los individuos, tomados en grupos pequeños aisladamente de la sociedad a gran escala, y luego aplicar los principios resultantes a todas las circunstancias posibles, incluyendo las que involucran millones de personas, instituciones políticas y económicas complicadas, y miles de años de historia. Lo más sorprendente es que él descubre en sí mismo intuiciones sobre los requisitos morales de las personas en un estado de naturaleza, que luego quiere adoptar como principios universales inalterados en sus efectos acumulativos, al aplicarlos a cualquier tipo de circunstancia.

Descrito abstractamente, este procedimiento parece desesperadamente equivocado.[3] Es difícil saber cómo alguien puede alcanzar seriamente opiniones morales firmes sobre los principios universales de la conducta humana sin considerar cómo

sería si fueran aplicados universalmente, en iteraciones que podrían traer complejos efectos a escala. Cuando pasamos de una descripción abstracta a una más sustantiva, se incrementa la inaceptabilidad de esta concepción. Pues la intuición que descubre Nozick en sí mismo es que todos tienen un derecho absoluto a ser libres de coacción, y un derecho absoluto a adquirir y disponer de su propiedad, siempre que no violen los mismos derechos de los demás, y siempre que su adquisición de propiedad no les dé, por ejemplo, el exclusivo título a la fuente de agua, antes pública, de una comunidad en el desierto.[4]

La intuición de Nozick es que cada persona tiene derecho a sus talentos y habilidades, y a todo lo que pueda hacer, obtener o comprar con sus propios esfuerzos, con la ayuda de los demás o por pura suerte. Tiene derecho a guardarlo o hacer lo que quiera con él, y a quien se lo dé tiene igualmente derechos sobre él. Más aún, cualquiera tiene derecho a lo que sea que resulte de la indefinida repetición de este proceso, aun con el paso de generaciones. Supongo que la mayoría de los lectores del libro de Nozick no encontrarán, en sí mismos, ningún eco de esta intuición y sentirán, en cambio, que no pueden formarse una opinión acerca de los principios universales de apropiación, adquisición y transferencia de propiedad, o de su existencia, sin considerar la significación de tales principios en su aplicación universal. Podríamos estar parcialmente de acuerdo con Nozick, respecto de lo que la gente debe hacer en las circunstancias limitadas que definen las relaciones interpersonales en el estado de naturaleza, pero no podemos estar de acuerdo en que la adecuada generalización de esos juicios es su aplicación inalterada a todos los casos, independientemente de cuán extensos y complejos sean. Se podrían basar, en cambio, en principios que dan estos resultados para transacciones en pequeña escala, pero con resultados muy diferentes para la especificación de condiciones generales de los derechos por ser aplicados en una escala indefinidamente amplia.[5]

El hecho es, sin embargo, que las intuiciones morales de Nozick parecen incorrectas, aun a pequeña escala. El niega que alguno de los derechos que detecta pueda ser violado para hacer un bien o para prevenir un mal. Pero, incluso si no es permisible asesinar o mutilar a un inocente a fin de promover algún resultado altamente deseable, los derechos protegidos no tienen todos el mismo grado de importancia. Se supone que

uno debe estar protegido, en orden de gravedad, de: ser matado, lastimado, dañado, sometido a fuerza, privado de libertad de muchos tipos diferentes (movimiento, asociación y actividad), de la destrucción o robo de su propiedad, o de la amenaza de alguno de los mencionados (con todas *sus* variaciones de gravedad). Es mucho menos admisible sostener que quitarle alguna propiedad a un hombre inocente es un medio no permisible para prevenir un gran mal, que sostener que matarlo no es permisible. Estos derechos varían en importancia, y algunos no son absolutos ni siquiera en el estado de naturaleza.

Las fuentes de la moral no son simples, sino múltiples. Por lo tanto, su desarrollo en la teoría política ha de reflejar esa multiplicidad. Los derechos limitan la obtención de fines valiosos, pero a veces pueden ser violados si los fines son lo suficientemente importantes. La única manera para progresar en la comprensión de la naturaleza de los derechos individuales es investigar sus fuentes y relaciones recíprocas y con los valores, cuya prosecución ellos limitan. Nozick dice poco sobre la base de la inviolabilidad de las personas, pero la siguiente observación indica hacia dónde él tendería a enfocarse:

> ¿[P]or qué no se puede violar el derecho de las personas por un bien social mayor? Cada uno de nosotros, individualmente, escogemos, algunas veces, padecer algún dolor o un sacrificio para obtener un beneficio mayor o para evitar un daño más grande: vamos al dentista para evitar, a la postre, un sufrimiento peor; hacemos algún trabajo desagradable por sus resultados; algunos se someten a dieta para mejorar su salud o su aspecto, algunos ahorran dinero para mantenerse en la vejez. En cada caso, algún costo es impuesto en razón de un bien mayor superior. ¿Por qué no sostener, de forma *similar*, que ciertas personas tengan que soportar algunos costos que beneficien a otras personas en razón de un bien social superior? Sin embargo, no existe ninguna *entidad social* con un bien que demande algún sacrificio para su propio beneficio. Hay sólo personas individuales, diferentes personas individuales, con sus propias vidas individuales. Usar a uno de estos individuos en beneficio de otros es usarlo a él y beneficiar a otros. Nada más. Lo que ocurre es que algo se le hace a él por el bien de otros. Hablar de un bien social superior encubre esta situación (¿intencionalmente?). Usar a una persona en esta forma no respeta ni toma en cuenta suficientemente el hecho de que es una persona independiente, que esta es la única vida que tiene. *El* no ob-

> tiene ningún bien compensatorio por su sacrificio y nadie está facultado a forzarle a esto, menos aún, el Estado o el gobierno que reclama su obediencia (mientras que a los otros individuos no) y que, por tanto, tiene que ser escrupulosamente *neutral* entre sus ciudadanos. (pp. 44-45, itálicas en el original, se omite nota al pie)

No está claro cómo piensa Nozick acerca de que los derechos individuales derivan del hecho de que la vida de cada persona es la única que tiene. Parece obtener la implicación de que un beneficio a una o más personas nunca puede compensar un costo pagado por otra persona. Esto, sin embargo, es una pretensión demasiado amplia, para los propósitos de Nozick. Es obviamente falsa e inadecuada como base para las restricciones en el trato de individuos.

Para que la compensación interpersonal tenga sentido no es necesario invocar la tonta idea de una entidad social, estableciendo con ello una analogía con la compensación intrapersonal. Lo que se necesita es la creencia, compartida por la mayoría de las personas, de que es mejor que diez personas reciban un beneficio antes que una sola lo reciba, que es peor que diez personas sufran un perjuicio antes que una sola lo sufra, es mejor que una sola persona se beneficie mucho antes que una sola se beneficie poco, etc. El hecho de que la vida de cada persona es la única que tiene no nos vuelve incapaces de hacer tales juicios, y si una elección entre tales alternativas *no* implica la violación de ningún derecho o título, sino sólo la distribución de tiempo o recursos limitados, entonces tomamos estas comparaciones como razones excelentes para elegir una alternativa, en lugar de la otra. Si podemos ayudar, tanto a diez personas como a una, que no está incluida en esas diez, y ayudamos a las diez, podemos decir que el rescatar a esos diez compensa la pérdida de esa persona, pese a que *ella* no obtiene ningún bien compensatorio por su sacrificio, y esta es la única vida que tiene.

De modo que, a fin de comparar posibles resultados de una acción, donde no está en cuestión la violación de derechos, es evidente que el carácter distintivo de los individuos no nos impide sopesar beneficios y perjuicios entre personas. Si intervienen restricciones especiales cuando tenemos que imponerle un sacrificio a alguien como un *medio* para obtener un resultado deseable, su fuente debe encontrarse en otro lado. Tales res-

tricciones no deberían derivarse de un principio que también tenga la consecuencia de que no se puede decir casi nada sobre la relativa deseabilidad de situaciones que involucren cantidades de personas diferentes.

Más aún, no podemos descubrir la fuente de los derechos de tipo general que invoca Nozick, concentrándonos, como él sugiere, en el significado de las vidas humanas individuales y el valor de diseñar y formarse una concepción general de la propia vida. Como todas sus sugerencias son vagas (véase pp. 58-59), todas sufren de un error de enfoque, porque se concentran sólo en rasgos de personas para las que es malo que les *ocurran* ciertas cosas, y que sea bueno para ellas tener la oportunidad de hacer ciertas cosas. Pero los derechos de la clase que le interesan a Nozick no son derechos a que ciertas cosas no nos *ocurran*, ni derechos a tener ciertas oportunidades. Se trata, más bien, de derechos a no ser deliberadamente tratado o usado de ciertas maneras, y a no ser deliberadamente interferido en ciertas actividades. Ellos dan lugar a reclamos, no en contra del mundo en general, sino sólo en contra de aquel que los viola deliberadamente. La *relación* entre el poseedor del derecho y el agente, antes que sólo la naturaleza intrínseca del poseedor y de su vida, debe formar parte del análisis del derecho y de la explicación de su fundamento.

Cualquier teoría de los derechos debe explicar este rasgo estructural, aun si no sigue a Nozick al elevar el ejercicio irrestricto de la voluntad como supremo principio de la moral. Es de fundamental importancia que mi derecho a no ser asaltado *no* sea un derecho a que todos hagan lo necesario para que yo no sea asaltado. Es meramente un derecho a no ser asaltado, y se correlaciona con el deber de los demás a no *asaltarme*. Esto no puede explicarse simplemente por el hecho de que es malo ser asaltado, que es simplemente un ítem en el catálogo de valores por el cual se tiene que sopesar la deseabilidad o no deseabilidad de sucesos o clases de sucesos. Que el asalto sea desagradable o malo no explica por qué su prohibición debería servir como restricción para la prosecución de otros valores, o la prevención de otros daños, aun si esos otros valores compensan la maldad del asalto en un puro cálculo de la relativa deseabilidad de posibles resultados. A veces, tenemos que elegir la alternativa menos deseable, porque, para lograr la más deseable, tendríamos que violar algún derecho.

Como señala Nozick (p. 41), las restricciones a la acción representada por derechos no pueden equivaler a la atribución de un gran disvalor a su violación, ya que eso haría permisible violar tal derecho, si, al hacerlo, uno pudiera evitar violaciones más numerosas o más serias del mismo derecho, por parte de otros. Esto no es verdad, en general. No es permisible, en opinión de Nozick (o en la mía), matar a una persona inocente, aun cuando fuera para prevenir el asesinato deliberado de otras tres personas inocentes. Un rasgo general de algo que puede legítimamente ser llamado un derecho es que su violación no sea traducible a una mera asignación de disvalor.

Por lo tanto, una explicación del fundamento de derechos tendría que concentrarse en el agente y su relación con la persona a la cual él está restringido de tratar de ciertas maneras, ni siquiera para lograr fines muy deseables. Y tendría que explorar la interacción entre esas restricciones y los fines que ellos restringen. No hay razón para pensar que, tanto en la vida personal como en la sociedad, la fuerza de cada derecho será absoluta o casi absoluta, esto es, imposible de ser revocada por consideraciones consecuencialistas. Los derechos a no ser deliberadamente matado, lastimado, torturado o encerrado son muy poderosos y limitan la prosecución de cualquier objetivo. Las restricciones más limitadas de libertad de acción en el uso de la propiedad, y restricciones acerca de contratos son, simplemente, menos serias, y por lo tanto, establecen restricciones menos poderosas.[6]

Más aún, hay una gran diferencia entre expropiar repentinamente la mitad de los ahorros de una persona, e imponer condiciones monetarias previas a actividades, gastos y ahorros, es decir, el ejercicio de impuestos habitual. El segundo es una forma mucho menos brutal de asalto a la persona.[7] El que esta clase de limitación de la libertad individual deba ser permitida, a fin de adquirir recursos para la promoción de fines deseables, es una función de la gravedad de la violación y de la deseabilidad de los fines. (Y, tal como observé, esto no quiere decir que esté justificado cuando el resultado es un equilibrio social máximo de beneficios y costos.)

Nozick respondería que tales fines pueden lograrse por medio de donaciones voluntarias, antes que por compulsión, y que la gente mejor situada y que lamenta la existencia de la pobreza debería donar porciones significativas de su haber para ayu-

dar a los menos favorecidos (pp. 257-260). Pero esto no es más aceptable viniendo de Nozick de lo que era cuando venía de Barry Goldwater. La mayoría de las personas no son generosas cuando se les pide que donen voluntariamente, y no es razonable pedirles que lo hagan. Es claro que hay casos en los que una persona debería hacer algo, aunque no sea correcto forzarla a hacerlo. Pero, en este caso, creo que lo contrario es correcto. A veces es adecuado forzar a la gente a que haga algo, aunque no sea cierto que lo deberían hacer sin estar forzados a ello. Es aceptable obligar a la gente a que contribuya para apoyar a los pobres por medio de impuestos automáticos, pero no es razonable insistir que, en ausencia de tal sistema, deberían contribuir voluntariamente. Esta última es una posición moral excesivamente exigente, porque reclama decisiones voluntarias que son bastante difíciles de tomar. La mayoría de la gente tolera un sistema universal de impuestos obligatorios, sin sentirse habilitados para quejarse, mientras que se sentirían justificados al rechazar un llamado a que contribuyan voluntariamente con la misma suma. Esto se debe, en parte, a la falta de seguridad de que los otros se comportarían de la misma manera, y al temor frente a las desventajas relativas, pero se trata también de un rechazo razonable a excesivas demandas voluntarias, que pueden ser más fastidiosas que las demandas automáticas al bolsillo.

Una teoría política que refleje estas complejidades morales debería asignarle a la sociedad la función de promover ciertos bienes y prevenir ciertos males, dentro de los límites establecidos por las diferentes restricciones de los diferentes derechos individuales. No debería juzgar procesos y procedimientos exclusivamente según su tendencia a producir ciertos resultados, ni tampoco juzgar resultados exclusivamente según los procesos que los produjeron. Las instituciones sociales y los procedimientos que las definen deberían ser evaluados tanto en referencia a su respeto a la libertad y los derechos individuales, como por su tendencia a promover fines deseables, como el bienestar general.

Nozick ofrece una clasificación de principios de justicia distributiva, en la cual dicha teoría no encuadra.[8] Después de definir un principio *histórico* como aquel que sostiene que la justicia de una distribución depende de cómo se produjo (p. 157) y un principio *de resultado final* como aquel que niega lo ante-

rior (p. 158), define un principio *pautado* como aquel que especifica "que la distribución debe variar de conformidad con alguna dimensión natural, con la suma de pesos de las dimensiones naturales de conformidad con un orden lexicográfico de dimensiones naturales" (p. 159).[9] Su propia teoría, la teoría del título, es fácil de describir en estos términos. Es un principio histórico no pautado, dado que especifica que cualquier distribución es justa si se alcanzó por medio de una serie de transacciones individuales entre personas con un título para hacerlo, por medio de los derechos naturales de adquisición y transferencia.

Pero supongamos una teoría que dice que una distribución es justa si resulta de un proceso gobernado por reglas que reflejan (a) la adecuación de ciertas pautas, (b) la deseabilidad de incrementar ciertos resultados buenos y disminuir ciertos males, independientemente de cualquier pauta, y (c) un respeto por los derechos individuales de diferente importancia. Tal teoría no sería en el fondo ni puramente histórica, ni puramente pautada. Sería formalmente histórica, pero el criterio de proceso, o "histórico", estaría determinado en parte por consideraciones de pauta y consideraciones de resultado final. Por lo tanto, el ataque de Nozick, concentrado en los principios pautados y los principios no históricos, no nos da ninguna razón para pensar que su alternativa sea correcta.[10]

Dejando de lado este defecto, su ataque sigue sin tener éxito. Nozick nos pide que imaginemos algún principio pautado realizado, y luego argumenta que su preservación exigiría interferir en la libertad individual: se le debería impedir a la gente que use como quiera lo que se les asignó bajo la pauta. Por ejemplo, la preservación de una distribución razonablemente equitativa exigiría que no se les permitiera a los individuos pagarle a Wilt Chamberlain 25 centavos por cada partido de baloncesto en el que pueden verlo jugar, con el supuesto de que él puede guardárselos, aun si representa un total de $ 250.000 por año. Esto es perfectamente obvio, y es parte de lo que se *significaría* con un principio de distribución pautado: la adopción de un sistema general de adquisición, impuestos e intercambio que tiende a preservar cierta pauta.

Esto parece representar un problema para Nozick y una posterior violación de libertad, sólo porque él, erróneamente, interpreta la noción de principio pautado como si especificara una distribución de *títulos absolutos* (como los que él defiende)

a la riqueza o propiedad distribuida. Pero el título absoluto a la propiedad no es lo que se le asignaría a la gente en una distribución parcialmente igualitaria. La posesión otorgaría la clase de derecho calificado que existe en un sistema en el que funcionan impuestos y otras condiciones para preservar ciertos rasgos de la distribución, permitiendo una elección, uso e intercambio de propiedad, que son compatibles con él. Lo que alguien tiene, en tal sistema, no será *su propiedad* en el sentido no calificado del sistema del título de Nozick. Pensar que es de otra manera da por sentado lo que hay que probar, y eso es exactamente lo que hace Nozick cuando dice:

> *No* hay duda de que si cada una de las personas tenía derecho al control sobre los recursos que tenían en D_1 [la distribución pautada original], puesto que esa fue la distribución (su distribución favorita) que (para los propósitos del argumento) dimos como aceptable... Si las personas tenían derecho a disponer de los recursos a los que tenían derecho (según D_1), ¿no incluía esto el estar facultado a dárselo, o intercambiarlo, con Wilt Chamberlain? (p.164, itálica en el original)

Este error quita fuerza al argumento.

Permítaseme volver brevemente al difícil tema de la igualdad. En tanto que la eliminación de la miseria, pobreza y enfermedad son probablemente fines sociales más importantes que el logro de la igualdad económica, este último es uno de los fines que algunas personas consideran legítimo. La opinión de Nozick es predecible. Si la desigualdad surge como resultado del libre ejercicio de títulos, no se la puede objetar en base a la injusticia, y de ningún modo se puede infringir la libertad para reducirla. Ya que las personas tienen derecho no sólo a la riqueza que heredan, sino también a sus dotes naturales (p. 221), serán justas las desigualdades posteriores que resulten de la utilización de estos.

Pero no hay ninguna razón para creer en un derecho natural absoluto a beneficiarse de las facultades naturales propias, aun si se reconoce algún derecho menos que absoluto a su libre ejercicio, como una limitación en la prosecución de igualdad o de otros fines sociales. Alguien que considere a la igualdad como un bien supondrá que su logro no tiene precedencia absoluta sobre la eficiencia, libertad y otros valores. Sin embargo, se requiere más que esto para responderle a Nozick, ya que no está

claro qué es lo que hace que la igualdad de distribución sea una cosa buena en sí misma. Nozick no reconoce el derecho del Estado a limitar la libertad para producir algún resultado meramente deseable. Pero, ¿por qué alguien, con una posición más estándar sobre derechos individuales, debería estar a favor de un fin de igualdad social y económica? Quizá puede argumentar que el nivel promedio de bienestar —tanto en términos materiales como en términos de satisfacción y autoestima— probablemente sería más alto en una sociedad relativamente equitativa que en otra que no lo es, en cuanto a la riqueza total comparable. Quizás argumente que los efectos políticos de la desigualdad económica son perjudiciales para la libertad individual y el bienestar general. Pero estas consideraciones, pese a ser muy importantes, no son razones para concebir la igualdad como un bien en sí mismo, aunque es una posición moral común.

No se puede defender la igualdad pretendiendo que las desigualdades son arbitrarias, salvo basándose en diferencias moralmente pertinentes entre personas. La arbitrariedad es un defecto moral sólo si puede ser contrastado con una alternativa elegida sobre la base de factores moralmente relevantes. Salvo que haya una justificación independiente de la igualdad, una distribución igual es tan arbitraria, desde un punto de vista moral, como cualquier otra. Para defender la igualdad como un bien en sí mismo, debería argumentarse que la mejora en el grupo de personas situadas más abajo en la escala de bienestar ha sido priorizada respecto de mayores mejoras de aquellas que están más alto, aun si estas últimas también afectaran a más personas. Pese a que esta posición resulta atractiva, no creo que haya sido nunca defendida en forma exitosa.

He tratado las limitaciones teóricas del libro. Hay mucho en él que no he analizado incluyendo el capítulo final, que describe una utopía pluralista libertaria (pp.287-319), y análisis interesantes de temas tan diversos como la teoría del valor del trabajo de Marx (pp. 246-255) y el trato de animales (pp.47-53). A diferencia de la mayoría de las obras filosóficas, el libro es divertido, de rápida lectura, y es un placer leerlo. El estilo de Nozick revela gran energía y alcanza un notable nivel de claridad y distinción. Es raro que uno dude de lo que Nozick defiende o de lo que uno niega cuando se rechaza su posición.

Notas

1. Nozick es el autor de tres artículos importantes: "Coercion" en *Philosophy, Science, and Method* (S. Morgenbesser, P. Suppes and M. White, comps., 1969); "Newcomb's Problem and Two Principles of Choice", en *Essays in Honor of Carl G. Hempel* (N. Rescher, Comp., Reidel, 1969); "Moral Complications and Moral Structures", *Natural Law Forum* 1 (1968). El libro alcanza su nivel de agudeza sólo en el cap. 4, "Prohibition, Compensation, and Risk" (pp. 54-88), un brillante análisis de la elección entre varios métodos de resolver una conducta injuriosa o peligrosa: cuándo prohibir, cuándo castigar, cuándo exigir compensación, cuándo compensar a alguien perjudicado por una prohibición. Se trata, además, del capítulo más importante para la teoría del derecho.

2. Esto ayuda a explicar el atractivo romántico del anarquismo. El intento de Nozick de refutar la posición anarquista de que incluso un estado mínimo viola derechos individuales no es, en mi opinión, exitoso. El sostiene in extenso que un Estado mínimo podría surgir por un proceso de mano invisible de un Estado de naturaleza, sin que el proceso viole ningún derecho de nadie: la gente podría reunirse en sociedades protectoras privadas, una de las cuales dominaría naturalmente un territorio, aunque no todos estuvieran de acuerdo en asociarse. Entonces podría ejercer control limitado, sin violar los derechos de nadie. Se supone que esto demuestra que un Estado mínimo es moralmente permisible. Pero, ¿por qué la mera idea de tal proceso habría de persuadir a un anarquista de esta conclusión? El anarquista ya estaría preparado de antemano a admitir que un Estado mínimo, establecido por acuerdo *unánime* por los participantes, sería permisible. El anarquista cree que ningún estado real puede ser de esta clase. Del mismo modo, puede reconocerle a Nozick haber imaginado otra forma en la que el Estado mínimo "podría" surgir, sin violar los derechos de nadie, aunque se basara en un acuerdo menos que unánime. Pero la probabilidad de que algún Estado real cumpliera con estas condiciones morales sería casi nula. El rechazo del anarquismo requiere el rechazo de sus premisas morales.

3. Nozick defiende el procedimiento en la sección denominada "Macro y Micro". Dice: [C]omplejos totales no son fáciles de analizar; no podemos fácilmente retener todo lo que es pertinente. La justicia de toda una sociedad puede depender de que satisfaga varios principios distintos. Estos principios, aunque individualmente obligatorios (atestigua esto su aplicación a un amplio rango de microcasos particulares), pueden producir resultados sorpresivos cuando todos se combinan... [N]o debemos depender de juicios sobre el todo como si proporcionaran el único o aun el cuerpo principal de información contra el cual verificar nuestros principios. El mejor camino para cambiar nuestros juicios intuitivos sobre un todo complejo es viendo las más grandes y a menudo sorprendentes implicaciones de principios fundados sólidamente en el micronivel (p. 203). Obviamente, aunque otro modo de cambiar juicios intuitivos sobre el alcance o verdad de principios en el micronivel es ver sus mayores implicaciones. El hecho de que los derechos de gobiernos deriven de los derechos de individuos no implica que podamos llegar a conocer los derechos de los individuos sin pensar en los gobiernos, así como el

hecho de que las propiedades de moléculas derivan de las propiedades de los átomos no implica que podamos llegar a conocer las propiedades de los átomos sin investigar las moléculas. Las conexiones lógicas y epistemológicas no tienen que ir en la misma dirección: aun si la filosofía política depende lógicamente de la ética, nuestro conocimiento de algunos aspectos de la ética puede provenir de una investigación en filosofía política.

4. Este último es la conocida cláusula de Locke en su teoría de adquisición de la propiedad pero, según Nozick, no funciona como una seria restricción en un sistema de libre mercado, cf. p. 182.

5. El ejemplo de apropiación que Nozick ofrece (p. 206) como respuesta a tal escepticismo —un derecho natural a no ser privado de los propios órganos vitales en beneficio de otros— es admisible, en parte, debido al carácter extremo de una agresión de este tipo y, en parte, porque no hay posibilidad de que la protección de este derecho nos conduzca a la acumulación de una gran riqueza hereditaria o a desigualdades de poder social y político.

6. El hecho de que pueda revocarse un derecho a fin de evitar consecuencias lo suficientemente serias, no quiere decir que se le pueda asignar a su violación un disvalor comparable al disvalor de esas consecuencias. Ya que ello le daría a la ocurrencia de dicha violación más peso del que tiene en realidad, en el cálculo de resultados (p.ej., cuando la cuestión es qué se debe hacer para prevenir tales violaciones por parte de otros). Por lo tanto, pese a que los derechos pueden ocasionalmente revocarse, la violación de los derechos de algunas personas no puede justificarse automáticamente, porque conduciría a una reducción de la más grave violación de los derechos de otras personas. Este tema surge en conexión con el arresto preventivo, escuchas telefónicas, búsqueda e incautación, que pueden ser útiles en la prevención de robos, asesinatos, asaltos y violaciones.

7. Se puede objetar que la aplicación de impuestos debe estar apoyada por la amenaza de una fuerza más potente o, incluso, de violencia. Pero esto surge sólo si hay resistencia en el primer nivel. Si la demanda original, no violenta, es legítima, puede tener lugar una escalada si el sujeto se resiste y usa métodos más violentos para resistirse a cada método sucesivo implementado para hacer valer el previo.

8. Naturalmente, la distribución es solamente una de las cosas que cualquier teoría política cubre pero, por el momento, podemos dejar a otras de lado.

9. "A cada uno según su necesidad" sería un ejemplo.

10. Más específicamente, sus argumentos en contra de Rawls se ven seriamente debilitados por un intento rígido de describir el principio rawlsiano de justicia distributiva como un principio ahistórico o de resultado final. Rawls no sostiene que la justicia de una distribución pueda determinarse independientemente de cómo se produjo. El cree que esta justicia depende de la justicia de las instituciones, incluyendo aquellas instituciones legales que definen apropiación y están involucradas en su producción. Estas se evalúan sólo parcialmente sobre la base de su tendencia a promover cierto estado distributivo final. Rawls, por ejemplo, le da prioridad a la preservación de la libertad individual, y aunque esto no significa lo mismo para él que para Nozick, ciertamente restringe los procedimientos por los que una distribución puede decidirse en forma justa. Véase, en general, J. Rawls, *A Theory of Justice* (Harvard University Press, 1971).

14

Hare: El pensamiento moral*

Este ensayo fue originalmente una reseña de Moral Thinking: Its Levels, Method, and Point, *de R.M. Hare (Oxford University Press, 1981). El compromiso de toda la vida de Hare por encontrar la fundamentación de la moral en el análisis del lenguaje es un ejemplo especialmente llamativo del error de tratar de entender un tema básico desde muy afuera, una tendencia alimentada por la idea, alguna vez prevaleciente, de que el lenguaje es la clave de todo en la filosofía.*

Según el profesor Hare, la mayoría de los filósofos morales están sumidos en la ignorancia. No pueden entender los claros principios del razonamiento moral que él estableció y desarrolló en sus dos extensos estudios sobre teoría ética, *The Language of Morals*[1] y *Freedom and Reason*[2], y que vuelve a explicar y desarrollar en la presente obra. Como un maestro de niños débiles mentales, no puede disimular su impaciencia al tener que repetir todo una y otra vez, pero la importancia de los temas le obliga a hacerlo.

El encuentra que el tema está infestado de "intuicionistas". Estos son los filósofos que limitan su pensamiento moral a un nivel: el nivel de los principios toscos, generales, *prima facie*, que fueron educados a seguir en la vida cotidiana, en la que hay que actuar sin tener mucho tiempo para pensar y sin un conocimiento perfecto de los hechos. Los intuicionistas toman estos principios, y los sentimientos morales asociados a ellos, como la base

* Reimpreso con autorización de *The London Review of Books*, 1-14 de julio de 1982.

del pensamiento moral, cuando en realidad derivan de otra cosa y pueden ser justificados o criticados sólo con moverse a otro nivel, donde no hay confianza en las intuiciones morales.

Este es el nivel de lo que Hare llama "pensamiento crítico", y sus reglas están completamente determinadas por la lógica de los conceptos morales: una lógica que se descubre en la investigación del uso linguístico y no en el ejercicio de alguna forma de intuición o juicio moral. Sólo al salir de la moral de este modo, según él, podemos esperar proporcionarle a esta una sólida fundamentación, en lugar de dar vueltas en círculos. Una vez hecho esto, descubriremos que la única posición moral posible de ser justificada sobre la base de la lógica de los conceptos morales es el utilitarismo. Y también resulta que el utilitarismo, cuando se lo interpreta adecuadamente, es lo que subyace a las intuiciones morales comunes, que los intuicionistas erróneamente consideran básicas y que usan para refutar el utilitarismo.

Hare ha llegado gradualmente a esta posición. Al principio, no creía que pudiera derivarse una única posición sustantiva de la lógica de los conceptos morales, y esto por buenas razones. Lo más inaceptable de esta posición es que implica que todos los desacuerdos morales fundamentales son, en cierto sentido, ilusorios. Suponiendo que las partes de un desacuerdo hagan afirmaciones morales, sólo pueden alcanzar conclusiones diferentes porque están en desacuerdo sobre hechos no morales, o porque al menos uno de ellos ha cometido un error lógico al aplicar los conceptos morales. No hay ningún desacuerdo que sea totalmente moral. Un libertario, que cree en la inviolabilidad de los derechos de propiedad individuales, y un socialista, que cree en promover la igualdad por medio de impuestos progresivos, descubrirán, al considerar cuidadosamente lo que quieren decir con las palabras "correcto", "erróneo", "debe" (*ought*) y "tiene que" (*must*), que ambos están comprometidos con un único método para decidir cuál posición es la correcta: un método empírico, que depende de cuál sea el sistema social que satisface en forma más eficaz las preferencias de todas las personas afectadas por él.

¿Cómo logra Hare sacar este gran conejo moral de algo que parece a primera vista una galera lingüística bastante pequeña y vacía? Su teoría de la lógica de los conceptos morales —el prescriptivismo universal— parece bastante inocente. Sostiene que un juicio sobre lo que debe hacerse en una ocasión particular nos compromete a prescribir que se haga lo mismo en

cualquier situación que sea similar en todas sus características universales, en particular, en cualquier situación en que las personas individuales involucradas sean diferentes o desempeñen diferentes roles.

Esto parecería ser algo sobre lo que el libertario y el socialista estarían de acuerdo. El libertario, que piensa que no debe pagar impuestos para proporcionar subsistencia a madres solteras indigentes, también piensa que, si él fuera una madre soltera indigente, o uno de sus hijos, estaría mal cobrarles impuestos a otros para que lo alimenten a él. Y el socialista quiere que los ricos paguen impuestos, independientemente de quiénes lo sean. Ambos pueden usar el mismo vocabulario moral para estar en desacuerdo acerca de lo que debe hacerse, porque difieren en cuanto a la importancia que se le debe asignar, universalmente, a la libertad, igualdad, propiedad, y a la satisfacción de las preferencias individuales.

Pero Hare ha llegado a creer que el prescriptivismo universal es mucho menos hospitalario respecto de las otras convicciones morales, porque lo interpreta de forma que le permita a cada persona prescribir universalmente sólo una cosa, dados los hechos de una situación particular, y que es la misma cosa que todos los demás deben prescribir. La única base racional de una prescripción universal es, según Hare, que nos ubiquemos en la posición de todos aquellos afectados por un acto determinado, preguntándonos si nos gustaría que se llevara o no a cabo si se tuvieran las mismas preferencias. Las preferencias que reconocemos que *tendríamos*, si estuviéramos en una situación dada, nos conducen inmediata e inevitablemente a preferencias tan fuertes como las que tenemos *ahora*, respecto de lo que *debería* pasar, si estuviéramos en esa situación. Entonces le agregamos todas estas preferencias hipotéticas, los pros y contras, tomando en cuenta sus diferentes pesos, y la elección de mayor peso decide a su favor si el acto debe o no llevarse a cabo, determinando si podemos prescribir universalmente que debería hacerse.

Se supone que todo esto se encuentra comprendido en la lógica de los términos morales utilizados por todos los hablantes de inglés y de otras lenguas traducibles al inglés. Pero esto, evidentemente, no es así. No sólo que mucha gente consideraría inválidas las críticas a sus opiniones morales basadas en este método, sino que incluso la mayoría de los que están de acuerdo con la posición moral de Hare no considerarían que los

que alcanzan juicios morales por otros métodos estén utilizando mal el lenguaje. En cambio, sostendrían que, si alguien cree, por ejemplo, que cada uno puede favorecerse a sí mismo por sobre los demás, sea o no que contribuya a la ventaja de todos a largo plazo, esa persona estaría en desacuerdo moral sustantivo con el utilitarismo en un nivel fundamental. Este es un desacuerdo no sobre lógica o sobre hechos, sino sobre el método para determinar lo que se debe prescribir universalmente.

En verdad no sorprende que Hare haya introducido de contrabando una intuición moral en su análisis del lenguaje moral, y esto es lo que le proporciona su método. Se trata de una intuición perfectamente respetable, la misma que Sidgwick consideraba la base del utilitarismo: "No debo preferir mi bien menor antes que el mayor bien de otro". Pero hay quienes no comparten esto. El paso de lo que uno querría para sí mismo, en cada circunstancia posible, a lo que uno querría que ocurriera universalmente no es un paso lógico, ya que hay más de una manera de combinar tales preferencias: la maximización simple de la satisfacción esperada, contando todas las posibilidades como igualmente probables, puede no ser la elección correcta, incluso para un individuo. Pero, más importante aún, el paso de lo que uno quiere que ocurra universalmente a lo que uno está preparado para prescribir que debe ocurrir universalmente no es un paso lógico. Un método alternativo para llegar a prescripciones universales es preguntar qué razones tendría cualquier persona para *hacer* en una situación dada. Una respuesta a esta pregunta es la utilitarista. Pero también hay otras respuestas, incluyendo aquellas que le permiten a una persona darles un peso extra a sus propias preocupaciones, y aquellas que incluyen sólo a algunos de los intereses de los demás entre sus preocupaciones.

Hare no puede extraer su teoría moral de la lógica de los conceptos morales, y la ilusión de que sí lo logra condena a la obra a un tipo de superficialidad, dado que le impide ver sus presupuestos más importantes, lo que lo priva de defenderlos. Este es un defecto del libro, aun si su posición moral sustantiva fuera correcta.

Pero su defensa del utilitarismo es parcialmente independiente de estas afirmaciones sobre el lenguaje moral. Hare presenta una versión de la posición que enfatiza la distinción entre las reglas ordinarias que deben seguirse en las circunstancias de la vida cotidiana, confusas y a menudo emocionalmente car-

gadas, y el método de pensamiento crítico, que puede ser utilizado en un momento sereno, sin limitaciones de tiempo e información, para trazar, evaluar o criticar esas reglas *prima facie* y resolver conflictos entre ellas. Sólo un ser omnisciente, cuya mente trabajara como un rayo, y que no padeciera de ninguna debilidad emocional, podría utilizar a pleno el método del pensamiento crítico utilitarista para cada acción que llevara a cabo. Los seres humanos reales tienen que prestar su conformidad al utilitarismo, siguiendo reglas que sean simples, claras, aprendibles, sin darle demasiado lugar a la decepción personal. Hare también piensa que ellas no deberían tener exigencias demasiado pesadas sobre nuestros motivos. Este es un punto curioso, al que desearía volver más adelante.

Hare dice que muchos de los argumentos más populares en contra del utilitarismo —los que afirman que exige alguna acción obviamente malvada como el asesinato, la tortura o el castigo de algún chivo expiatorio inocente, en circunstancias suficientemente extravagantes— se originan por ignorar esta penetrante distinción entre los niveles crítico e intuitivo del pensamiento moral. Creo que tiene razón. Una forma razonable de utilitarismo abogará por la creación de instituciones estables que promuevan el bienestar general pero, también, que inculquen virtudes sólidas, hábitos de conducta, reacciones instintivas y sentimientos que lleven a la gente a elegir lo mejor, en general, en las circunstancias ordinarias de ignorancia y tentación. No conocemos los resultados totales de muchos de nuestros actos con suficiente detalle como para permitir que el cálculo utilitarista proceda acto por acto. Es mucho mejor que haya un rechazo contundente de la tortura, cualquiera sea la razón, y no que la gente se sienta preparada para evaluar su posibilidad caso por caso. La justificación utilitarista de entrenar a la gente a tener estas inhibiciones morales depende de los hechos reales de la vida humana, de modo que no es un argumento en contra del utilitarismo decir que las intuiciones firmes evitarían que alguien hiciera lo mejor sobre la base del conocimiento perfecto en un tipo de situación tan extravagante que raramente o nunca ocurre. Creo que es muy improbable que la forma de utilitarismo de Hare justifique, en la práctica, atrocidades morales obvias (aunque atrocidades morales hayan sido a menudo defendidas por razonamientos utilitaristas espurios). La cuestión teóricamente difícil es si el enfoque de dos niveles de Hare nos da una explicación adecuada de los

derechos, deberes y prohibiciones de la moral común, o si estos tienen fuentes independientes, sin derivar completamente de la intuición única de imparcialidad moral que promueve el utilitarismo.

Tengo dos dudas: primero, si la posición de Hare nos da la explicación correcta de las intuiciones morales comunes, en los casos en que las apoya; segundo, si el utilitarismo realmente coincide con la moral común, tanto como Hare cree. Al considerar requisitos tales como veracidad, lealtad o fidelidad respecto de nuestros compromisos con los demás —requisitos que exigen que ignoremos la utilidad general aparente en casos particulares—, Hare nos pide que creamos que no tienen fuentes independientes, ni razones que resultan de las circunstancias particulares o de las relaciones en las que se aplican. En otras palabras, si, sin tomar en cuenta la utilidad, me reprimo de mentirle a alguien, o de engañar a alguien que confía en mí, la justificación última descansa en la utilidad general de una disposición bastante inflexible a no hacer esas cosas. Pero, incluso si dicha disposición tiene utilidad general, sucede otra cosa más. Es inaceptable decir que el rechazo intuitivo al engaño personal sea sólo un artificio de educación en base a principios garantizados por la utilidad. La posición de Hare nos hace considerar el sentido de inmediatez que tienen estas exigencias como si fuera una clase de ilusión, y a todos los conflictos morales que parecen surgir de ellas como susceptibles de ser resueltos con sólo elevarse al nivel de pensamiento utilitarista crítico. Pero, sin el apoyo de su axioma que dice que el único método de la justificación moral es la apelación a la imparcialidad pura, el diagnóstico de ilusión no es convincente, y parece probable la intervención de otras razones, más personales, de un tipo básico en la moral ordinaria.

En realidad, las razones personales parecen hacer que la moral ordinaria se desvíe sustancialmente incluso de un utilitarismo de dos niveles. Esta es mi segunda duda respecto de la posición de Hare. El suaviza el carácter exigente de la imparcialidad utilitarista mediante un recurso ilegítimo que facilita la explicación de otros aspectos de la moral común en términos de su teoría. Este recurso consiste en adaptar los principios *prima facie*, que se pide a la gente que siga en la vida cotidiana, a la parcialidad natural de la gente común, hacia sí mismos y sus allegados. Un utilitarista elogiará a los individuos por su virtud supererogatoria de hacer más, pero no le verá el

sentido de exigirles a los demás, o a sí mismo, un nivel de beneficencia que sería simplemente rechazado por ser demasiado oneroso. Como dice Hare: "Si yo fuera tan piadoso como Albert Schweitzer o la Madre Teresa, podría tener principios mucho más exigentes que los que tengo de hecho".

Yo lo diría de otro modo: si Hare fuera muy piadoso, podría ser más utilitarista. Tal como él es, está comprometido con principios *prima facie* para la conducta de vida que a menudo le permiten hacer cosas que, para los estándares utilitaristas, serían incorrectas, simplemente porque, como el resto de los no piadosos, no está dispuesto a digerir los principios más estrictamente imparciales, cuya aceptabilidad-utilidad, si *fueran* aceptados, sería mayor. Si él no piensa que está actuando mal al comportarse en forma menos abnegada que Albert Schweitzer, entonces le está dando a su parcialidad natural un veto sobre las exigencias excesivas de imparcialidad, lo cual significa salirse del utilitarismo pleno, aun cuando considere diferencias de nivel.

Yo encuentro la posición de Hare más aceptable que el utilitarismo puro pero, para poder resolver el asunto, habría que ahondar en los fundamentos de la ética. Estos no se encuentran solamente en el análisis de los conceptos morales, aunque estoy de acuerdo con Hare en que tampoco se encuentran en la sistematización de nuestros sentimientos morales. Mi opinión es que los métodos de pensamiento moral de los cuales disponemos actualmente son bastante primitivos; que, mientras captamos los tipos básicos de razones morales, incluyendo aquellas sobre las que el utilitarismo está exclusivamente basado, no tenemos un método adecuado para combinarlos de modo de obtener un resultado cuando hay conflictos entre ellos, como sucede constantemente en la vida real, cuando diferentes reclamos atañen a una decisión. No sé cómo elegir entre las diferentes interpretaciones de la igualdad moral, ni creo que la teoría ética actual haya agotado las posibilidades. Hare ofrece un método único, y su libro lo presenta con la lucidez, economía y atractivo que hace su obra tan agradable de leer. Pero, a pesar de que vale la pena estudiarlo atentamente, parece confirmar la idea de que, en la ética, la simplicidad teórica no es una virtud.

Notas

1. Oxford University Press, 1952.
2. Oxford University Press, 1963.

15

Hare: Los fundamentos de la imparcialidad*

Este trabajo, que es más bien conciso, trata algunos de los puntos mencionados en el ensayo anterior, con mayor detalle. El volumen en el que fue publicado incluye las respuestas de Hare a cada una de mis críticas.

I

¿Puede una teoría ética ser demasiado exitosa? Si propone una fundamentación de la argumentación moral que vaya más allá del alcance del desacuerdo moral, puede dejar sin explicar la existencia e intratabilidad de los desacuerdos morales reales. Si trata de apelar sólo a hechos fuera de la moral, que todas las partes deben admitir, entonces es probable que los desacuerdos morales más fundamentales resurjan, ya sea en la interpretación de lo que ha sido demostrado, o en la disputa acerca de los supuestos morales ocultos en las premisas.

Creo que el intento de Hare de fundamentar la moral en la lógica de los conceptos morales cae en este problema. Es verdad que cuando la gente no está de acuerdo acerca de lo correcto o incorrecto, y de cómo deben vivir, quieren significar la misma cosa con los términos que utilizan para expresar sus diferencias. Pero esta base común no proporciona por sí misma

* Reimpreso con autorización de *Hare and Critics*, Douglas Seanor y N. Fotion (comps.) (Oxford University Press, 1988).

los materiales necesarios para resolver sus desacuerdos, y el intento por mostrar que lo hacen, inevitablemente introduce elementos moralmente controvertidos en la interpretación de la supuesta base común. Esto ocurre en otras ramas de la filosofía, cuando se critica una u otra posición por violar la lógica de los términos con que se expresa. El error, si es que hay alguno, normalmente se encuentra en otra parte.

Puede ocurrir a veces, en el caso de disputas especializadas, casi legalistas, entre miembros de la misma secta moral, que el método para dirimir un desacuerdo moral pueda extraerse de los significados que las partes del desacuerdo deben usar en común. Pero creo que esto no es posible respecto de aquellos temas morales de gran escala, que poseen el mayor interés filosófico. En tal caso debemos ir más allá de la terminología para encontrar que no hay ningún método común disponible que no sea en sí mismo un tema de controversia moral. Lo mejor que podemos esperar es descubrir argumentos y contraargumentos que expongan las bases del desacuerdo en forma más clara dentro de los confines de la moralidad y, quizás, identificar algunas diferencias no morales intervinientes. Esto nos sitúa en una mejor posición para continuar con la búsqueda de un método sobre el cual podamos asegurar un acuerdo, a medida que emergemos de la Edad del Bronce moral.

La teoría desarrollada en *Moral Thinking*[1] sostiene que el contenido posible de una moral se encuentra más estrechamente restringido por la lógica del lenguaje moral que Hare defendió en la época de *The Language of Morals.*[2] Esto se debe a que él incluye en su versión posterior del significado de los juicios morales una interpretación de su universalidad que representa un fuerte requisito sustantivo de imparcialidad entre todas las personas. Es esta interpretación lo que me gustaría analizar, centrándome sobre todo en su relación con la lógica del lenguaje moral, más que en su aceptación independiente como suposición moral.

Hare cree que los desacuerdos morales reales, si no están basados en una confusión, pueden ser mejor explicados como diferencias acerca de lo que se deriva de dicho requisito, en conjunción con los hechos (diferencias debidas, en parte, a la variación en la corrección del razonamiento y, en parte, a la variación en creencias sobre los hechos no morales pertinentes). Pero creo que algunos desacuerdos se sitúan en el nivel del

principio mismo: desacuerdos en el seno de la moral acerca del rigor y la interpretación adecuada de la imparcialidad moral.

II

¿Qué es lo que nos lleva desde la mera universalidad a la forma rigurosa de imparcialidad de Hare? La mera universalidad de las afirmaciones morales es relativamente incontrovertida, y podría ser considerada como parte de su significado. Si formulo un juicio sobre lo que yo u otra persona debería hacer, estoy comprometido con la posición de que cualquier otro en las mismas circunstancias (analizadas tan sutilmente como se quiera), debería hacer lo mismo. El juicio particular es una consecuencia de principios más generales, quizá sumamente generales. Pero esto, por sí solo, no lleva a nada parecido a la imparcialidad en el contenido de los juicios morales; más aún, nos dice muy poco sobre su contenido. El egoísmo ético, por ejemplo, satisface perfectamente la condición de mera universalidad. Y Hare, en su análisis anterior sobre el "fanático",[3] reconoce que algunas posiciones espantosas podrían ser adoptadas, en cierto sentido, como prescripciones universales.

Obtenemos imparcialidad sólo si le damos una respuesta particular a la pregunta: ¿Cuál es la actitud que tomo respecto de los actos abarcados por un principio universal tal como este, cuando juzgo que ellos deberían (o no) hacerse? Esto es, necesitamos una explicación de qué es *prescribir* que algo se haga, que todos actúen de determinada manera. La posición de Hare es que una prescripción de que algo debe hacerse es la expresión de un deseo o preferencia de que se haga. Y una prescripción universal expresa un deseo de que lo mismo se haga en todos los casos similares en el mundo real y también en todos los casos hipotéticos similares que difieran de los reales sólo en las identidades de los participantes.

Si intercambiamos las identidades de todas las personas en el mundo con cada posición en todos los casos reales abarcados por una prescripción universal dada, tendremos un gran número de casos. Pero Hare supone que se puede considerar mi preferencia, con respecto a esta universalización, como dependiendo de lo que quisiera que me ocurra a mí mismo, en aque-

llos casos en que yo ocupo alguna de las posiciones. El sostiene, además, que lo que quiero que me ocurra a mí mismo en cada uno de esos casos dependerá directamente de lo que querría la persona que realmente ocupa la posición, dadas sus preferencias y dada una información completa. Y, finalmente, Hare sostiene que lo que yo quiero que se haga universalmente, en todos estos casos, estará determinado por un equilibrio de las preferencias que tengo en cada uno de ellos, de modo que se maximiza la satisfacción total de preferencias.

Con esto, llegamos al utilitarismo de preferencias: lo que es correcto hacer es aquello que maximiza la satisfacción total de las preferencias de todas las partes afectadas. (Dejo de lado aquí las complicaciones que resultan de la distinción entre los niveles crítico e intuitivo del pensamiento moral. Mi análisis tiene que ver exclusivamente con lo que Hare llama "nivel crítico".)

Aquí nos encontramos con una rigurosa posición moral sustantiva, que implica una cantidad de elementos moralmente controvertidos que no pueden ser atribuidos a la lógica del lenguaje moral. Específicamente, incluye una posición sobre cómo los juicios morales deben proporcionar una consideración igual a todos, cómo deben entenderse los intereses de todos para los propósitos morales, y cómo los intereses en conflicto de diferentes individuos deben combinarse para alcanzar conclusiones morales.

Podemos exponer el camino del prescriptivismo universal al utilitarismo, por medio de una serie de proposiciones:

1. Decir que se debe hacer algo moralmente es prescribir que se haga eso en todos los casos hipotéticos similares con las identidades intercambiadas.
2. Prescribir que se haga algo es expresar el deseo de que eso se haga.
3. Lo que deseamos que se haga en una serie de casos hipotéticos con las identidades intercambiadas depende sólo de lo que deseamos que se haga en aquellos casos en que ocupamos alguna de las posiciones.
4. Lo que deseamos que se haga en un caso hipotético en el cual estamos en el lugar de otro, con sus deseos y preferencias, depende de lo que creemos que esa persona haría en esas circunstancias.

5. Lo que deseamos que se haga universalmente en una serie de casos hipotéticos es una función simplemente aditiva de lo que deseamos que se haga en cada uno de ellos.

Las proposiciones (2), (4) y (5) son controvertidas y, en el contexto, moralmente sustanciales. Me parece particularmente poco convincente la afirmación[4] de Hare de que algo como (4) es una verdad conceptual. Pero quiero concentrarme en (2), y, en menor medida, en (5). En tanto que (4) asegura que la teoría moral resultante tomará la satisfacción de preferencias como la medida básica de valor, (2) y (5) aseguran que será una forma de consecuencialismo agencialmente neutral y maximizador. La corrección de las acciones dependerá de su tendencia a aumentar la cantidad de valor que es igual para todos los agentes y que es definida independientemente de la perspectiva del agente. Si Hare tuviera razón en que todo esto es parte del significado de los términos morales, entonces las alternativas más importantes al utilitarismo no podrían ni siquiera expresarse coherentemente en el lenguaje moral.

III

Podemos empezar por analizar (2), preguntando qué interpretaciones alternativas sobre lo que es prescribir un acto estamos dejando de lado. La respuesta de Hare a esta pregunta es que lo que debemos prescribir universalmente deriva de su definición de lo que es prescribir universalmente. Para poner la respuesta en tela de juicio, es necesario proponer una definición alternativa, no necesariamente una incompatible con la posición moral sustancial de Hare, sino una que no la implique. Creo que lo mejor es partir de la idea de que todos tienen una razón para hacerla.

La premisa (2) tiene una consecuencia peculiar. Cuando yo prescribo universalmente en el sentido definido por esta premisa, esto no implica que crea que todos tienen una razón para actuar de acuerdo con mi prescripción: en el mejor de los casos, implica que reconozco una razón para que yo la obedezca (aunque eso se deduce solamente si suponemos que al usar el lenguaje moral reconozco una razón por la que me importa la universalización). Pero un juicio moral debe tener una conexión

con la razón práctica, no sólo para la persona que lo produce, sino también para las personas a las que se dirige. Si se supone que un juicio moral es correcto, este nos hace suponer que los otros tienen una razón para obedecerlo, aun si ellos todavía no la reconocen como tal.

Esta afirmación puede ser expresada con un juicio moral, aun si no expresa el *deseo* del que lo emitió de que los demás actúen como él dice que deben actuar. Es sólo cuando el juicio se refiere a su propia conducta que tal deseo debe incluirse.

No se supone que esto dé cuenta completamente del juicio moral. Debería decirse algo más para diferenciar las prescripciones morales de la clase de los juicios de racionalidad práctica en general; es decir, las razones morales deben tener cierto grado de independencia de los intereses particulares y deseos del agente, por ejemplo. Pero, independientemente de lo que habría que agregar, la condición de que debe implicar la existencia de razones por parte de la persona a la que se dirige asegura la conexión con la motivación, la que debería preservarse en cualquier enfoque sobre la prescripción moral. Se deduce que si yo me aplico una prescripción moral a mí mismo, debo reconocer una razón para actuar. Esta conexión con la motivación es una condición esencial para que un juicio sea prescriptivo. Pero no es esencial que, además, cualquier motivación o deseo de mi parte se exprese con una prescripción de lo que otra persona deba hacer, aun cuando algunas concepciones morales particulares establezcan esta conexión.

La posición de Hare se basa en una interpretación de los juicios morales como imperativos en un sentido bastante literal, y no en el sentido en que Kant los concebía como imperativos. En sentido literal, en la vida diaria, yo no le diría a alguien que haga algo, a no ser que yo desee que lo haga. Pero, en el sentido kantiano, la enunciación, sea de un imperativo hipotético o categórico, se conecta en forma más cercana a la concepción del hablante sobre las razones que motivarían a su interlocutor, que a los deseos del hablante. Esta concepción consideraría mi juicio sobre alguna otra persona como prescriptivo, aun si yo no tuviera ningún deseo de que el otro lo siga, siempre que *él* no pueda aceptarlo sinceramente sin aceptar alguna razón para acatarlo.[5] En su interpretación de las prescripciones universales de la moral, Hare no da lugar a esta versión. Pero me parece natural y perfectamente inteligible

concebir una prescripción universal simplemente como una afirmación sobre cómo todos deben comportarse, tal que si alguien está sinceramente de acuerdo con ella, debe reconocer una razón para actuar de esa misma manera.

Esta interpretación es bastante débil, ya que no dice nada acerca de qué tipo de razón debe ser. Podría ser del tipo que el utilitarismo reconoce, pero también puede no serlo. La pregunta es cómo los intereses de todas las partes afectadas por el acto, incluyendo al agente, entran en la generación de las razones universales que el juicio moral implica.

Una respuesta a esta pregunta es que los intereses de todas las partes afectadas generan razones sobre una misma base, en proporción al peso de sus preferencias individuales. Esto es equivalente a la posición de Hare del contenido de la moral, ya que corresponde a lo que uno quisiera, adoptando el punto de vista de cada una de las partes, uno por uno, y combinando todas estas perspectivas vicarias, de algún modo, en una preferencia colectiva. Pero es concebible que haya otros sistemas de razones universales, que también consideren a todos, en algún sentido.

Específicamente, algunas posiciones morales sólo se pueden formular en términos de razones en las relaciones de parentesco. Una posición bastante común, sea correcta o no, es que un individuo puede considerar sus propios intereses y los de aquellos a los que está vinculado, con más fuerza que los de personas extrañas: no hay una única medida neutral del bien que se requiere moralmente que todos traten de promover (aunque tal valor neutral pueda formar *parte* de la base de la moral). Otros principios en las relaciones de parentesco, se encuentran en la creencia en restricciones deontológicas del tipo que limitan lo que hagamos intencionalmente, como opuesto a lo que podemos causar. A fin de evitar las complicaciones al analizar una licencia moral, que es la negación de una exigencia moral, permítaseme usar el ejemplo de una exigencia en las relaciones de parentesco, de favorecer los intereses de los hijos propios por sobre los intereses de extraños, un requisito que mucha gente aceptaría. Por el momento, no me interesa su verdad, sino su carácter formal.

Salvo que tal posición moral derive en forma disimulada de algo más universal, evidentemente no puede expresar una prescripción universal de acuerdo con la premisa (2). Si alguien

cree que debe favorecer a sus propios hijos, cree que cualquiera debe hacer lo mismo, y que esto es verdad para todos los casos similares reales o hipotéticos, incluyendo aquellos en que él ocupa una posición diferente. Pero esto no quiere decir que *desee* que todos deben actuar de este modo en todos los casos. Si él se ubica en cada posición diferente sucesivamente, y considera lo que le gustaría que se hiciera, y con qué rigor, podría llegar a un resultado muy diferente.

Por ejemplo, un norteamericano o europeo adinerado podría referirse a sus obligaciones para con sus hijos para explicar por qué él no debe (aun si tuviera la voluntad de hacerlo) contribuir con la mitad de su ingreso para mitigar el hambre, porque, aun cuando esto salve a muchos niños africanos de desnutrición, disminuiría drásticamente el nivel de vida de sus propios hijos. Pero, si se pusiera en el lugar de alguno de los niños africanos desnutridos, que podría salvarse de la muerte por su sacrificio, o en el lugar de sus padres, su deseo de donar superará con creces el deseo de no donar, tal como ocurre en su situación real. El podría, en realidad, pensar que, si estuviera en su posición, estaría justificado en tomar por la fuerza los recursos necesarios, si pudiera hacerlo. De modo que esta obligación relativa en las relaciones de parentesco sostiene que cada persona tiene una razón para preocuparse por sus hijos, que puede, en casos de conflicto, llevarlo a querer que otros prioricen los intereses de sus hijos tan tenazmente como lo harían si le dieran a esa razón todo su peso. Y el resultado puede ser que lo que uno prescribe universalmente no es lo que uno desea que deba hacerse universalmente.

La respuesta de Hare es que una obligación especial de ese tipo aparece sólo en el nivel del pensamiento moral intuitivo: tal vez se lo pueda justificar en el nivel crítico, apelando a la utilidad general de que la gente tenga una disposición profundamente arraigada a ejercer un cuidado especial con sus hijos; pero si no puede ser justificada de este modo, no existe esa obligación especial.[6]

Quizás Hare tenga razón. Pero, si alguien está en desacuerdo y sostiene que hay muchas más obligaciones y licencias especiales de las que el utilitarismo podría permitir, no puede ser refutado apelando a la lógica de los conceptos morales que ambos comparten, ya que la afirmación lógica crucial de Hare, que prescribir universalmente es expresar un deseo universal

de una clase particular, incluye la posición moral central que está en discusión.

Es difícil saber cómo resolver estas disputas, si no es por medio de la lógica. Su tema real es la motivación humana, específicamente, la razón práctica. No tenemos un método general para descubrir lo que la gente tiene razones para hacer y, por lo tanto, para descubrir esos principios universales de la razón práctica, referidos a la manera en que debemos considerarnos mutuamente, que forman la base de la moral.

Derek Parfit argumenta que algunos principios en las relaciones de parentesco, como el que he analizado antes, pueden ser criticados porque son colectivamente autofrustrantes. Por ejemplo, puede darse el caso de que, si cada uno de nosotros favorece a los propios hijos, los resultados sean peores para todos ellos, que si ninguno lo hace. Puede incluso darse el caso de que cada uno de nosotros pueda beneficiar personalmente a sus propios hijos menos de lo que podría si nadie sigue esta política.[7] Pero, en el mundo real, es evidente que, si nadie favoreciera a sus propios hijos, o si todos redujeran el grado de favoritismo, algunos niños ganarían y otros perderían; esta política no sería autofrustrante, en sentido estricto.

De modo que la crítica de tales principios no puede ser meramente interna. Debe ser una defensa de la imparcialidad radical entre personas, como una fundamentación motivacional de la moral. La pregunta es: ¿Tengo alguna razón para regular mi conducta con principios diseñados para darles igual peso a las preferencias de todos?

Para Hare, no se trata de una pregunta dentro de la moral, ya que cree que se deduce de la lógica de los conceptos morales que estas son las únicas clases de principios que pueden ser correctas. El argumento moral se detiene en este punto y, si hay más argumentos, deben ser de otra clase. Hasta cierto punto, Hare cree que es posible apelar a razones de prudencia al plantear la cuestión motivacional.[8] Pero, si bien hasta cierto punto dichas razones recomiendan utilitarismo, es probable que funcionen todavía mejor, para la mayoría de nosotros, en defensa de una moral menos exigente con licencias para las relaciones de parentesco que nos favorezcan a nosotros y nuestros allegados.

Al final, nos quedamos con una confrontación entre alguien que está dispuesto a hacer sólo lo que puede prescribir univer-

salmente, en el sentido estricto implicado por la premisa (2), y alguien, que Hare llama amoral, quien no prescribe nada en ese sentido y, por lo tanto, no puede usar el vocabulario moral. Como creo que esto es una disputa dentro de la moral, describiría como amoral sólo a alguien que esté más afuera, que se niegue a hacer prescripciones universales de un tipo mucho más débil: no tiene opiniones sobre cómo la gente debería comportarse, ni razones para actuar que todos deberían reconocer. Alguien que no piense que debe hacer lo que desea que otros hagan en casos similares puede no ser amoral, sino alguien con una posición moral diferente.

La cuestión de si los principios en las relaciones de parentesco tienen un lugar en la moral en el nivel fundamental es compleja en muchos aspectos. Incluye el análisis de los derechos y licencias tanto como de las obligaciones especiales. La posición de Hare sobre todas estas excepciones aparentes a la moral neutral de los agentes es que ellas pueden explicarse, en tanto sean válidas, como consecuencias en el nivel intuitivo del pensamiento utilitarista en el nivel crítico. Pero la cuestión teórica central es si el método utilitarista es correcto en el nivel crítico, y esto depende, creo, de si un consecuencialismo imparcial tiene un peso dominante como principio regulador de la razón práctica. Esta es, probablemente, la primera cuestión de la teoría ética. Requiere que consideremos, además, los grados alternativos de subordinación de nuestra conducta en la revisión de un punto de vista externo, independiente, que abstraiga quiénes somos. No creo que dispongamos de un método general para pensar sobre estos tipos de cuestiones.

IV

Permítaseme ir brevemente a otro tema. Pasaré sobre la premisa (4), salvo para mencionar que la cuestión de si nuestra preocupación por los otros deba estar determinada por sus preferencias, o por alguna otra cosa —una restringida subclase de sus preferencias, o su interés medido por algún estándar objetivo—, ha merecido extenso análisis recientemente.[9] Una vez más, lo que Hare supone aquí no es, por cierto, una tautología moral. Pero lo que quiero analizar ahora es cómo intereses en

conflicto deben combinarse en la deliberación moral, una vez reconocidos, en la forma que sea.

Hare responde a esta pregunta con la premisa (5). Lo que debo hacer está determinado al equilibrar los pesos de las preferencias que yo tendría para diferentes cursos de acción, si estuviera en la posición de cada una de las partes afectadas, incluyendo la del agente. El curso de acción con el mayor peso cuantitativo de preferencias favorables sobre las no favorables, para todas estas posibilidades en conjunto, es el correcto. Es el que yo prescribiría si tuviera que hacerlo universalmente.

Mi pregunta no es si esto es lo que yo *desearía*, dada esta elección, sino si esta forma de equilibrio es la forma correcta para darles a todos la misma consideración moral. Obviamente, se trata de *una* manera de acordar consideración igual: las preferencias de cada uno (o intereses, si se usa otra medida de beneficio o daño individual) se cuentan del mismo modo en el total. Pero, incluso si suponemos que la moral, o por lo menos un componente de ella, nos exige que seamos imparciales, ¿es esta la manera de hacerlo?

El problema sobre un principio maximizador es conocido. ¿Pueden las preferencias relativamente débiles, o los intereses menores de un gran número de personas, legítimamente superar las preferencias o intereses mucho más fuertes de un número pequeño de ellas? Esta clase de agregación es particularmente incómoda, cuando la mayoría está significativamente mejor situada que la minoría, en otros aspectos. Parece posible que la respuesta maximizadora más rigurosa esté equivocada en algunos casos, es decir, que una ventaja marginal a la mayoría de los mejor ubicados no debería ganar siempre.

Hare, como Rawls, encuentra el modelo de razonamiento moral en una situación especialmente limitada de elección racional individual. En lugar de ignorar quién es quién, y muchas cosas más, Hare usa el instrumento de la identificación imaginativa con cada persona, para garantizar que el punto de vista de cada una de ellas será considerado en su totalidad. Pero, la pregunta es si podemos combinar nuestra preocupación por el bienestar de diferentes personas con los mismos principios que son adecuados para resolver conflictos de intereses de una sola persona. La posición de Hare es clara: "No puedo ver la razón para no adoptar aquí la misma solución que

adoptamos en los casos en que nuestras propias preferencias están en conflicto entre sí".[10]

Esta cuestión ha sido ampliamente analizada, con resultados no concluyentes, en años recientes. Yo mismo pienso que hay un problema en la posición que sostiene que la compensación entre vidas no puede tener la misma forma que la compensación dentro de cada vida, y que los mismos principios no pueden gobernar las dos clases de conflicto. Si tenemos el mismo interés en los intereses de cada uno, nuestra preocupación por toda esa gente estará tan dividida como lo está su objetivo, y no es obvio cómo se pueden combinar esos componentes.

Me siento inclinado a pensar que debemos darles prioridad a aquellos individuos que están más necesitados de ayuda, porque son objeto de una preocupación que tiene prioridad respecto de nuestras preocupaciones por los mejor situados, aun si estos son los más numerosos. Cuando no hay compensación interpersonal de beneficios y daños, parece razonable algún tipo de jerarquización de prioridades, al menos como parte de la solución. Pero, no hay duda de que hay otras alternativas mejores, que no he imaginado. La teoría moral sufre aquí un empobrecimiento de posibilidades. Ninguna de las opciones conocidas parece claramente adecuada. Reitero, la situación requiere que busquemos un método de pensamiento que no tenemos todavía, en vez de extraer uno de las reglas que ya gobiernan nuestro lenguaje moral.

V

Mi objetivo limitado ha sido mostrar que los elementos de la lógica de los conceptos morales de Hare, que ocupan un papel central en su argumento a favor del utilitarismo, se entienden mejor, no como una parte de la lógica de esos conceptos, sino como premisas morales, cuya fundamentación aún no ha sido descubierta y que puede ser rechazada sin violar la lógica del lenguaje moral. Los fundamentos de la moral todavía no han sido descubiertos, tampoco en ningún otro lugar. Pero Hare, al hacer que sus principales tesis morales sean una parte de la definición de la moral, deja la búsqueda de su fundamento fuera de la teoría moral, que es el ámbito al que pertenece.

Hay otro inconveniente en su método, que tiene que ver con la motivación moral. Hare a veces compara su posición con la

de Kant. Pero, en la interpretación de Hare, los imperativos de la moral son hipotéticos. Alguien tendrá razones para obedecerlos sólo si le importa actuar del modo que podría prescribir universalmente. Podemos mostrarle que tiene razones de autointerés para que esto le importe, es decir, ser el tipo de persona que se regirá por principios morales. Pero, salvo que algún motivo externo como aquel esté presente, no tendrá ninguna razón para ser moral. No tendrá ninguna razón para preguntarse si podría prescribir universalmente la clase de conducta que se va a llevar a cabo, y ninguna razón para sentirse influido por la respuesta a esta pregunta, si es que la plantea. (Si él concluye que no podría prescribir universalmente esta clase de conducta, podría decir simplemente "¿Y qué?") Si alguien no tiene los motivos que, de acuerdo con Hare, están supuestos en el uso del lenguaje moral, entonces la teoría de Hare implica que el argumento moral no puede revelar ninguna razón que no reconoce todavía.

Sólo si comenzamos con una concepción de la moral como una clase de afirmaciones sobre cómo todos tenemos razones para comportarnos, vamos a descubrir, al final de nuestra explicación, que la cuestión de la motivación ha sido dejada fuera de la teoría ética, para ser resuelta, quizás, con una dudosa apelación al autointerés. La búsqueda de los fundamentos de la moral debería ser parte de una teoría general de las razones para la acción. Cualquier argumento que se ofrezca en este contexto en apoyo de una posición moral, debe mostrar que tenemos razones para actuar de acuerdo con ella. Una teoría que se apoya en una fuerte suposición motivacional dentro de las condiciones para usar el lenguaje moral, y que luego deriva el contenido de la moral a partir de la lógica de este lenguaje, no ha encontrado un fundamento seguro de la moral que defiende. Las duras y viejas preguntas han sido meramente excluidas del campo del argumento moral.

Notas

1. Oxford University Press, 1981 [*El lenguaje de la moral*].
2. Oxford University Press, 1952.
3. *Freedom and Reason* (Oxford University Press, 1963), cap. 9.
4. *Moral Thinking*, pp. 95-96.

5. No me refiero aquí a la posición de Kant sobre el contenido del imperativo categórico que, naturalmente, no incluye ninguna referencia a lo que podamos desear que todos hagan. Me refiero al *concepto* de un imperativo categórico.

6. Cf. *Moral Thinking,* p. 199.

7. Véase Derek Parfit, *Reasons and Persons* (Oxford University Press, 1984), pp. 100-108.

8. *Moral Thinking,* cap. 11.

9. Por ejemplo, Rawls, en su teoría de los bienes primarios, y Scanlon, en "Preference and Urgency", *Journal of Philosophy* 19 (1975).

10. *Moral Thinking*, pp. 109-110.

16

Williams: Un pensamiento de más*

Este ensayo fue originalmente una reseña del libro Moral Luck: Philosophical Papers 1973-1980 *(Cambridge University Press, 1981), de Bernard Williams. La maravillosa expresión, "suerte moral", de Williams, es una de las invenciones lingüísticas que expanden notablemente nuestro conocimiento. Fue el título de una ponencia que desarrolló en una de las reuniones de la Sociedad Aristotélica en 1976, en la que participé como discutidor. Le robé el título para mi respuesta, que fue reeditada más tarde en* Mortal Questions, *en 1979 [Trad. cast.* La muerte en cuestión*]. Pese a que el origen de la frase estaba especificado en una nota al pie, lamento escuchar que algunas personas creen que yo la inventé, lo cual es un error que tal vez me moleste más a mí que a Williams, cuya modestia siempre me facilita hacer las reseñas de sus libros.*

Hace más de una década, Bernard Williams atacó la objetividad de la ética con creciente intensidad. El niega que la ética pueda decirnos la verdad sobre lo que debemos hacer o cómo debemos vivir, mediante el descubrimiento de principios que se revelan cuando miramos nuestras vidas y sociedades, no desde un punto de vista subjetivo, sino imparcial e independiente. Tal punto de vista debería ser el análogo, en el campo de la elección y acción, de la objetividad buscada por la ciencia al abstraerse de cada percepción individual del mundo. La abstracción científica nos permite formarnos una concepción de cómo es el mundo en sí mismo, antes de cómo lo vemos desde

* Reimpreso con autorización de *The Times Literary Supplement*, 7 de mayo de 1982.

cada lugar o, incluso, de cómo generalmente se nos aparece a criaturas como nosotros.

El análogo mayor sería la posición de que los juicios morales son juicios sobre una realidad moral externa, así como lo es la astronomía sobre una realidad física externa, pero Williams niega mucho más que esto. Niega que pueda haber principios de conducta universales que deban recomendarse internamente, incluso a todos los seres humanos racionales, por más diferentes que sean sus sentimientos personales, ambiciones y vínculos. El lo niega, porque la ética, si tiene que referirse realmente a la acción, debe involucrar a la voluntad, y él cree que la motivación comienza y se limita a un conjunto de deseos y sentimientos que varía mucho de una persona a otra, e, incluso, en una persona de un momento a otro. Este problema teórico de la objetividad de la ética se conecta íntimamente con la cuestión más inmediata de cuán completa puede ser la sumisión que la moral nos demanda. Williams está dispuesto a oponerse con gran resistencia. Rechaza tanto la posición utilitarista de que debemos vivir de una manera tal que promueva el bienestar de cada uno imparcialmente, como la posición kantiana de que debemos vivir de acuerdo con principios que podrían ser adoptados imparcialmente como reglas que cada uno debe seguir. Y tiene serias dudas sobre cuán viable o importante es la moral, en cualquiera de sus formas.

Su menosprecio hacia la visión *sub specie aeternitatis* ya se encuentra en su contribución en *Utilitarianism: For and Against*, publicada en 1973 (donde su coautor, J. J. C. Smart, estaba *a favor* del utilitarismo). Desde entonces, sus trece ensayos compilados en *Moral Luck* conforman un libro sumamente interesante, producto de una fina inteligencia filosófica y de un temperamento irreverente, concentrado en problemas difíciles e inevitables. Williams escribe en forma vívida y con fuerza, y logra tratar cuestiones abstractas junto con problemas de la vida real de una manera que resalta sus conexiones. Algunos de los ensayos son cortos o fueron escritos rápidamente, pero, en su conjunto, forman una colección importante. A pesar de que su tesis central es, en mi opinión, falsa, cualquier persona interesada en el estado actual de la filosofía moral debe prestarle atención a este libro.

Mientras que la mayoría de los ensayos se refieren a la ética, dos de los mejores tratan la objetividad y el realismo fuera

de la ética, un tema que también recibe un profundo tratamiento en su libro *Descartes: The Project of Pure Enquiry* (1978). Williams ha estado en general preocupado por el tema de cuán fuera de nosotros mismos podemos o debemos tratar de llegar, en el pensamiento y en la acción. Pero su tratamiento de la moral no es la consecuencia de un escepticismo más amplio: Williams no se encuentra entre el número incomprensiblemente grande de filósofos que niegan la objetividad del conocimiento fáctico, o científico. Sus ataques a lo que considera pretensiones exageradas en ética son específicos y dependen de su afirmación de desempeñar un papel en la vida individual.

El impulso trascendente en la teoría ética aparece tanto en sus afirmaciones sobre el contenido universal de la moral, como en sus afirmaciones de autoridad suprema sobre nuestra vida. Williams desafía a ambas afirmaciones (sea en su forma utilitarista o kantiana), pero no está seguro de cómo dividir su oposición entre las dos. Un desafío en relación con el contenido diría que cualquier moral debe basarse en las actitudes y sentimientos dominantes de la persona a la que se aplica la moral y, por lo tanto, no puede ser impersonal. Un desafío respecto de la autoridad diría que, dado que la moral se basa en sólo algunos de nuestros motivos, debería a veces ser anulada por otros, que para nosotros son más importantes: no sólo motivos de autointerés, sino de altruismo hacia ciertas personas, o nuestro compromiso a lograr o perseguir algún proyecto u objetivo especial. (Naturalmente, el término "debería" no expresa aquí un juicio moral.)

La ambivalencia aparece, por ejemplo, en el ensayo que da título a la colección, en el cual se trata la cuestión de si las acciones pueden ser evaluadas retroactivamente, en virtud de resultados que no fueron ni podrían haber sido previstos con certeza, con anterioridad. ¿Puede la suerte determinar si lo que hicimos está justificado o no, y si lo está, si se trata de una cuestión interna a la moral, o una limitación a ella? Williams argumenta que no siempre es posible actuar de modo tal que uno no tendrá ninguna razón para reprocharse a sí mismo, cualquiera sea el resultado: "La perspectiva de elección deliberativa sobre la vida propia es constitutivamente *desde aquí*. Correlativamente, la perspectiva de evaluación con mayor conocimiento es necesariamente *desde allí*, y no sólo no puedo garantizar qué es lo que ocurrirá de hecho, sino que, en

última instancia, no puedo garantizar desde qué perspectiva de evaluación será mi arrepentimiento mayor y más fundamental". Si no hay ningún punto de vista de evaluación externo que sea válido atemporalmente, no puedo ni siquiera pretender que mi elección actual sea objetivamente correcta respecto de mis *creencias* habituales sobre la probabilidad de diversos resultados.

Williams analiza el ejemplo de Gauguin, o alguien como él, quien abandona a su familia para dedicarse al arte, y que podrá considerar retrospectivamente su elección como justificada si, y sólo si, resulta ser un gran artista. "El proyecto, en interés del cual se toma la decisión, es aquel con el que el agente se identifica, de modo tal que, si logra llevarlo a cabo, su punto de vista de evaluación será desde una vida que obtiene gran parte de su significación para él, a partir de ese mismo hecho; si no lo logra, puede, necesariamente, no tener esa significación en su vida."

Esto parece implicar relatividad respecto del punto de vista en el *contenido* de los juicios morales. Pero, ¿son morales estas evaluaciones? Williams dice que el hecho de que Gauguin no pueda usarlas para justificarse a sí mismo ante su familia no muestra que no lo sean. En general, cree que la justificación moral de la acción de una persona no tiene que silenciar las quejas de los demás si fueron agraviados por ella. (Esto lo retoma en el ensayo "Politics and Moral Character", que trata el problema de las manos sucias.) Pero aquí Williams se siente atraído por la idea diferente de que, a pesar de que Gauguin no está moralmente justificado por su éxito, su suerte moral consiste, no obstante, de la relación entre su vida y la moral: él podría, por razones morales, lamentarse por lo que hizo si hubiera fracasado, pero esas razones morales están empantanadas para él, para nosotros y también para su familia, por todos sus cuadros. Y esto depende de su suerte de convertirse en un gran artista. Williams concluye que si nosotros rechazamos, por estas razones, la autoridad de cualquier moral que pretenda decirnos lo que sería correcto independientemente de los resultados, nos quedaríamos con una moral menos importante de lo que nuestra moral se considera habitualmente. Yo mismo preferiría decir que Gauguin estaba justificado por su éxito, pero que esta justificación, a pesar de que probablemente no sea moral, no fue tampoco un mero producto de su punto de vista retrospectivo, sino que era algo más objetivo.

Williams ataca también el predominio de principios impersonales, en el conocido ensayo "Persons, Character, and Morality", del cual mencionaré dos puntos. El sostiene, en contra de Charles Fried, que es un error querer una *justificación* para salvar la vida propia antes que salvar a un extraño de ahogarse, en lugar de tirar una moneda, en un caso de elección si/no, y objeta que la persona que rescata a la otra piensa que es *permisible* salvar a su esposa, por lo que tiene "un pensamiento de más". Williams afirma que puede haber ciertos proyectos básicos en la vida de una persona, que son una condición para que ella quiera seguir viviendo, y que no puede ser razonable pretender que abandone esos proyectos en el nombre de la moral imparcial —sea utilitarista o kantiana— presumiblemente porque, si se va, la moral no puede atraparlo.

Esta es una forma audaz y más bien original de chantaje: "¡antes que tener que servir al mejor bien para el mayor número, o el imperativo categórico, prefiero morir!" Sería una afirmación que poca gente haría sin engañar. En tanto que las exigencias de algunas formas de moral impersonal sobre la motivación individual pueden parecer excesivas, es mucho más difícil que lo que Williams hace, simplemente negarle jurisdicción al punto de vista impersonal respecto de nuestros sentimientos y compromisos más fundamentales.

El hombre que se arroja a las olas para salvar a su esposa no tiene argumentos kantianos o utilitaristas de regla dándole vueltas en la cabeza, pero esto no le impedirá que, retrospectivamente, tenga algo para decir, aunque sea a sí mismo, que le justifica no haber hecho nada para inhibir el impulso natural de parcialidad extrema (supongamos que tenía mejores chances de salvar a un extraño). El tendría el mismo pensamiento en el caso en que la situación se invierta y su esposa se ahogue, porque el único nadador que podía rescatarla era otro hombre que, en cambio, rescató a su propia esposa. Tal vez Williams diría que, en este caso, cualquier sentimiento que no sea la pena sería indecente, pero yo creo que la mayoría de nosotros tenemos el impulso a tratar de conducir nuestra vida personal y formarnos nuestros objetivos básicos, de tal forma que puedan estar de acuerdo con un punto de vista impersonal desde el cual todos son juzgados del mismo modo.

Se trata de uno de nuestros puntos de vista, después de todo: es una parte importante de algo que nos hace humanos. Y, si la

parcialidad debe tener límites, estos deben ponerse desde afuera. La dificultad consiste en lograr algún tipo de integridad en la vida humana sin agobiar su esfera personal con una imparcialidad arrolladora, ni destruir el punto de vista impersonal en el nombre de lo que uno, personalmente, tiene que hacer. Ambas reacciones a los problemas que Williams plantea tan vívidamente tienen el sabor de reacciones hacia uno mismo: por un lado, la culpa respecto de lo personal y egoísta; por el otro, la rebelión contra las amonestaciones impersonales de la conciencia. El descubrimiento de una alternativa, con la que podamos vivir, considero que es la tarea de la teoría ética.

Williams cree que no puede haber tal cosa como una teoría ética, ya que nada tan supuestamente universal puede explicar por qué derecho se legislan los sentimientos morales. En "Razones internas y externas", desarrolla la posición de que todas las razones que una persona puede tener para hacer o querer hacer algo deben derivar del razonamiento deliberativo de su acervo motivacional subjetivo, que consiste simplemente en aquel dado a una persona en cierto tiempo, aunque pueda cambiar. Si hubiera tal cosa como razones externas, que valen para todas las personas racionales independientemente de los detalles de su acervo motivacional, entonces deberían ser alcanzables, al punto de convertirse en motivacionalmente eficaces, por medio de algún proceso racional diferente al del método deliberativo habitual de derivar conclusiones prácticas a partir de motivos existentes. "No veo ninguna razón para suponer", dice Williams, "que estas condiciones puedan ser satisfechas."

No veo ninguna razón para suponer lo contrario. Williams mismo está preparado para admitir que "hay indeterminación esencial en lo que puede contar como un proceso racional deliberativo". Y piensa que razonar puede, a veces, producir cambios en el acervo motivacional. ¿Por qué no puede haber, tal como sostuvieron diferentes defensores de la moral imparcial, alguna forma de percepción sobre nuestro lugar, que no es único en el mundo, que nos lleve a reconocer que debemos vivir de determinada forma que podamos adoptar desde afuera, y también desde adentro, para todos los que están situados similarmente. Además, habría que preguntarse por la forma que este resultado debe adoptar y cuán dominante es. Pero, ¿por qué ponerse en el lugar del otro no es una forma genuina de razonamiento?

Permitirle a la perspectiva imparcial tal autoridad regulativa no tiene que implicar necesariamente el dejar de lado o la supresión de esos sentimientos, afectos, proyectos e individualidad de carácter que Williams tan correctamente enfatiza. En "Utilitarismo y autoindulgencia moral", dice que, incluso si el fin del conocimiento objetivo es disociar el pensamiento acerca del mundo de lo que uno distintivamente es, y quizás, de cualquier cosa distintivamente humana, "que no puede ser el objeto del pensamiento y experiencia moral, que debe involucrar principalmente captar al mundo de modo que uno pueda, como un ser humano particular, vivir en él". Esto es cierto, pero me parece que establece la tarea central de la teoría ética, en vez de dar una razón para abandonarla.

El mismo ensayo trata en forma sutil y convincente la crítica de que aquellos que rechazan ciertos medios para el logro de las mejores consecuencias generales deben ser acusados de autoindulgencia moral. Williams reconoce que puede existir esta autoindulgencia, pero se encuentra ante una preocupación reflexiva para con la propia virtud, respecto de la que el utilitarista puede ser tan culpable como cualquier otro. Aunque Williams no siente ninguna simpatía por esto, tiene gran respeto hacia el fenómeno menos autoconsciente del *carácter* moral, que hace imposible o difícil o necesario que una persona haga ciertas cosas, aunque no pueda dar razones que proporcionen una justificación universal. Esto es importante para las virtudes morales, y es una de las razones por las que los conflictos de valor no siempre pueden ser resueltos al reducirlos a una base racional común y sumarles los componentes.

En "Conflictos de valor", Williams defiende el tema relacionado de que, en un contexto de sentimiento moral compartido, el hecho de que no es necesaria ninguna razón adicional para apoyar una distinción moral (entre, por ejemplo, aborto e infanticidio) no quiere decir que esa distinción sea irracional. " 'No puedes matarlo, es un niño' es más convincente, como razón, que cualquier otra que pueda ofrecerse para que sea una razón." Esto parece correcto, pero también parece compatible con críticas o justificaciones posteriores.

"La verdad en el relativismo" sostiene que no podemos valorar moralmente como correctos o incorrectos sistemas de creencias morales que estén en una confrontación puramente "nocional", en tanto opuesta a "real", con la nuestra: es decir,

sistemas que, como incompatibles con los nuestros, no representan opciones reales. Williams da como ejemplos la vida y perspectiva de un jefe de la Edad del Bronce griega, o de un samurai medieval, sosteniendo que "carecen de la relación con nuestras preocupaciones, que por sí misma le da sentido y sustancia a la valoración". Se trata de un relativismo sumamente restringido, pero la referencia a "sentido y sustancia" no le otorga ningún argumento independiente a su favor. Si hubiera tal cosa como la objetividad moral, contrariamente a lo que piensa Williams, entonces tal vez no habría mucho sentido, pero por cierto habría sustancia en la valoración moral de la esclavitud faraónica, aun cuando no sea una opción real para nosotros.

No puedo hacer justicia respecto de los ensayos finales, "Wittgenstein y el idealismo" y "Otro tiempo, otro lugar, otra persona". El primero es el mejor tratamiento que conozco de la cuestión de hasta qué punto el escrito posterior de Wittgenstein expresa un idealismo trascendental de la primera persona plural, usando *nosotros*, que es el plural que desciende del *yo* idealista. El mundo, entonces, no es *mi* mundo, sino *nuestro* mundo. Williams encuentra esta posición en el segundo Wittgenstein, junto con otro problema, también presente en el *Tractatus*, de cómo expresarlo sin hablar sin sentido (diciendo, por ejemplo, que las verdades de la matemática dependen de nuestras decisiones).

El último ensayo demuestra en forma brillante la incoherencia del positivismo y sus herederos que proviene de tratar de unir el empirismo verificacionista, con su énfasis en la perspectiva humana, con la cosmovisión científica, que busca una representación del mundo, que no es desde aquí ni de ninguna perspectiva particular. Nótese que esto es una imagen especular del ataque a la fusión kantiana de la independencia impersonal con el fin moral básico de guiar la vida individual dentro del mundo.

Moral Luck es un libro atrapante. Williams es un escritor provocativo y estimulante, y es el principal detractor contemporáneo de las ambiciones trascendentes en filosofía moral. Al atacar estas ambiciones, y al dudar sobre la importancia preeminente de la moral, Williams se sitúa en el corazón mismo del tema. Lo que importa es si está en lo cierto.

17

Williams: Resistiendo a la teoría ética*

Este ensayo fue una reseña del libro Ethics and the Limits of Philosophy, *de Bernard Williams (Harvard University Press, 1985). Los ataques escépticos de Williams hacia la moral impersonal no han tenido un impacto extraordinario en este campo: en general, Williams provoca a quienes no están de acuerdo con él a que escriban mucho sobre la teoría ética insurgente del tipo que él cree imposible. Sin él, la vida nos sería a los kantianos mucho menos interesante.*

Este libro tan absorbente es un ataque a la posibilidad de la teoría ética, en tanto disciplina. Su tesis central es que la visión *sub specie aeternitatis*, a pesar de ser un punto ventajoso para buscar una comprensión del mundo natural, es totalmente inadecuada como punto de vista para construir nuestro mundo, el mundo en que debemos vivir. Ya que la pregunta de cómo debemos vivir en relación mutua es la pregunta decisiva de la ética, se deduce que la típica ambición de la teoría ética —descubrir principios morales universales, objetivos— no es realizable. Esto no se debe a que no haya verdades éticas, sino a que la verdad en este ámbito no es ni universal, ni objetiva, ni basada en la razón. Es sólo otra falacia de los filósofos que, sin una fundamentación filosófica, el pensamiento ético es imposible.

* Reimpreso con autorización de *The Journal of Philosophy* 83 (1986).

Algunos de los argumentos a favor de esta conclusión son francamente filosóficos, especialmente los negativos, que tienen que ver con la naturaleza y límites de la objetividad; entre los representantes recientes del nuevo romanticismo, este libro se destaca por el nivel de su argumentación. Pero también se hace evidente una corriente de exasperación respecto de la teoría ética moderna, que considera que es de poca utilidad para tratar los temas prácticos de la vida contemporánea y que está más alejada de lo concreto que la religión, a la que sustituye. Acerca de la esperanza de alcanzar resultados prácticos por medio de una justificación filosófica de la moral, Williams dice, en forma bastante desalentadora, "¿Qué puede lograr la justificación del profesor, cuando le tiran la puerta abajo, le rompen sus anteojos, y se lo llevan?" (p. 23)

Williams ataca a los candidatos principales a favor de una teoría general, y ofrece la esperanza de una forma alternativa de pensamiento ético, más escéptica y menos ambiciosa, pero también más realista social y psicológicamente, que sus rivales: él cree que podemos conservar muchos de los ideales de la Ilustración sin una teoría del conocimiento moral de la Ilustración. La concepción positiva no es más que una esperanza y no está totalmente desarrollada, pero creo que sería un error que los especialistas en teoría ética abandonen sus esfuerzos en virtud de argumentos negativos. Sin embargo, es un gran libro, que brilla con inteligencia y estilo y, como cualquier otra obra de interés fuertemente filosófico, del que se puede aprender mucho sin, por ello, estar convencido de su contenido.

Williams establece un contraste entre objetividad científica y razón práctica. Toma la posición, por lo común poco razonable, de que la ciencia nos dice algo sobre el mundo que ya está allí, independientemente de nuestras teorías, y la elabora en términos de la idea (tomada de su libro anterior sobre Descartes) de una concepción "absoluta" de la realidad, una concepción que es, en un grado máximo, independiente de nuestras perspectivas humanas o personales y de sus peculiaridades, pero, al mismo tiempo, coloca esas perspectivas dentro del mundo, y explica por qué el mundo se les aparece de esta manera y por qué podemos formarnos tal concepción absoluta. El modo como el mundo realmente es explica cómo podemos formarnos una concepción correcta de él, al trascender las apariencias, aun si partimos de puntos de vista diferentes.

Pero el pensamiento ético se refiere a la cuestión de cómo debemos vivir, no cómo son las cosas. Aun si esta pregunta admitiera la convergencia de diferentes perspectivas en una sola respuesta, la explicación no podría parecerse en nada a la que ofrece la ciencia, que es la convergencia en nuestras creencias sobre cómo es el mundo. Las convicciones y desacuerdos éticos no son como percepciones causadas por la acción en nuestros sentidos de un único mundo externo.

De modo que la defensa de principios objetivos en ética requiere una explicación alternativa de cómo es posible la convergencia sobre la verdad, y una teoría del error que explique los fracasos evidentes y frecuentes de dicha convergencia. Pero Williams defiende que no existe tal alternativa. La matemática no proporciona un modelo mejor para la verdad y el conocimiento ético, que el que nos ofrece la ciencia: toda explicación debe adecuarse específicamente a su tema. La cuestión es si alguno de los candidatos principales tiene alguna posibilidad de estar en lo cierto.

No toda teoría ética trata de proporcionarle a la ética una fundamentación universal e ineludible. Williams critica ciertas formas de utilitarismo y contractualismo, representadas por R. M. Hare, John Rawls y Thomas Scanlon, quienes parten del supuesto motivacional de alguna forma de imparcialidad, sin aspirar a que dicha imparcialidad, en sí misma, sea universalmente necesaria. El sostiene que estas posiciones son, de diferentes modos, motivacionalmente poco realistas. (Sobre el estado peculiar de conciencia moral requerido por el utilitarismo indirecto o de dos niveles, por ejemplo, dice: "No se puede separar, excepto mediante una disociación impuesta e ilusoria, al yo de la persona que teoriza del yo sobre cuyas disposiciones ella teoriza" [p. 110]). Pero quisiera concentrarme en sus objeciones hacia las dos tradiciones de teoría estrictamente fundacional que él considera que hay que descartar: la kantiana y la aristotélica.

Una trata de fundamentar la ética en nuestra racionalidad, la otra, en nuestra humanidad. Esto quiere decir que la primera tiene pretensiones, de algún modo, más universales que la segunda, pero ambas tratan de probar la autoridad de la ética, basándola, inevitablemente, en algo que nos es esencial. Y ambas pretenden ofrecer una solución objetiva, en la forma de la adopción de ciertos principios y disposiciones, desde un pun-

to de vista exterior a las contingencias culturales y personales de nuestra vida; un punto de vista que podría ser compartido con seres muy diferentes, histórica y personalmente, de nosotros mismos.

Según la posición kantiana, tal como la interpreta Williams, los requisitos éticos emergen de la idea de libertad racional, una idea que se aplica tanto a la justificación de creencias como a la formación de deseos e intenciones. En ambos casos, se supone que nosotros podemos alejarnos de nuestras primeras impresiones o inclinaciones y decidir, desde este punto de vista externo, si debemos admitirlas, en cuanto consideraciones racionales. Así como yo debo preguntar, ante la evidencia, si soy libremente racional, qué debe creer esta persona que soy yo, del mismo modo debo preguntar, a la luz de las consideraciones a las que tengo acceso, qué debe hacer esta persona que soy yo. En ambos casos, la respuesta dependerá no sólo de mi situación e inclinaciones personales, sino de principios que se deberían aplicar a todos los que están similarmente situados; y, quizás, por último de principios de pensamiento o de razón práctica que son totalmente generales.

Podemos dejar de lado aquí los resultados sustantivos supuestamente derivados de este paso. Williams objeta el paso en sí mismo. Su argumento en contra es breve. Cree que la analogía entre la reflexión fáctica y práctica es pobre, porque la deliberación fáctica no es esencialmente de primera persona. Al tratar de decidir qué hacer, yo simplemente trato de decidir qué es verdadero, y las relaciones objetivas de la evidencia a la conclusión deberían dominar el resultado, sea que la evidencia provenga de mi experiencia o de la de otros. Pero:

> Es muy diferente para la deliberación de la acción. La deliberación práctica es, en todos los casos, de primera persona, y la primera persona no es reemplazable derivativamente o naturalmente por *nadie*... El *yo* que se aleja de mis deseos en la reflexión racional es todavía el *yo* que tiene esos deseos, y va a actuar, empírica y concretamente; y no se convierte, al alejarse en la reflexión, en un ser cuyo interés fundamental sea la armonía de todos los intereses. (pp. 68-69)

Si este fuera un argumento eficaz, debería funcionar contra todas las versiones de la idea de razonamiento práctico desde

un punto de vista independiente, sea que le den autoridad a la armonía de todos los intereses, o no. El argumento es que, al alejarme de mis deseos, no traigo a la cuestión ningún punto de vista motivacional que los supere en autoridad. Mis creencias perceptivas, naturalmente, se someten a tal autoridad, porque están orientadas hacia el mundo externo a mí, pero mis deseos prerreflexivos se orientan sólo a algo más personal: a lo que voy a hacer. La reflexión, si ha de tener algún efecto práctico, no puede dejar de lado a este *yo* y sus motivos particulares, no puede ponerlo como mero *objeto* de consideración.

La alternativa aristotélica —con la que Williams simpatiza mucho más, debido a su énfasis en las virtudes y en el carácter total de la vida— basaría una respuesta universal, a la pregunta de cómo deberíamos vivir, en una teoría de la naturaleza y del bienestar humanos. Si hubiera una naturaleza humana universal y si, contemplándola desde afuera, viéramos que hay una única forma de vida individual y social y un conjunto de disposiciones emocionales y prácticas que fueran lo mejor para todo ser dotado de esa naturaleza, tendríamos una base objetiva para adoptar (por parte de los humanos) esa vida y esas disposiciones. Se armonizaría la visión interna y externa, y la ética derivaría de una concepción verdadera de nuestro lugar en el universo.

La objeción de Williams a esto no es que sea ininteligible, sino simplemente que ya no es más aceptable creer que exista tal teleología controladora de la naturaleza humana. Aparte de la base biológica, el bien humano está subdeterminado por la naturaleza humana. Más aún, hemos aprendido a esperar que las disposiciones que definen cualquier forma de vida más particular serán, vistas desde afuera, menos convincentes.

> El punto de vista externo de sus disposiciones es accesible para el agente. Pero, si este, al reflexionar, trata de abstraerse totalmente de aquellas disposiciones y de pensarse a sí mismo y al mundo como si no las tuviera, entonces no debería sorprenderse si no puede tener un cuadro adecuado del valor de algo, incluso de sus propias disposiciones. No puede, justamente porque esas disposiciones son parte del contenido de su yo actual. (p. 51)

Desde afuera, es evidente que muchas perspectivas incompatibles son compatibles con la naturaleza humana; el último

soporte de un punto de vista ético puede sólo encontrarse en las disposiciones reales del agente, no santificado por una teleología universal.

¿Con qué nos quedamos entonces? Williams cree que hay verdades acerca de cómo debemos vivir, y que pueden ser conocidas; pero sostiene que esto sólo tiene sentido si adoptamos una concepción no objetivista del asunto. Si tratamos de ascender hacia un punto de vista objetivo, perderemos las perspectivas que hacen posible el conocimiento ético. Si insistimos en movernos sólo con lo que es, para nosotros, una *consideración* racional, nos meteremos en un rincón vacío desde el cual nada parece importarnos. Debemos comenzar con lo que somos: no meramente lo que somos en tanto seres humanos o racionales, sino lo que somos como criaturas particulares de una cultura particular.

Una de las señales de esta particularidad es la capacidad de aplicar conceptos sustantivos o "gruesos", como *cobarde, mentira, brutalidad, gratitud,* etc. Estos términos pueden usarse solamente desde una determinada perspectiva ética, por aquellos que tienen disposiciones adecuadas para sentir y actuar. Sin embargo, pueden usarse para emitir enunciados verdaderos y expresar conocimiento ético, conocimiento que puede no ser accesible a alguien que no puede ocupar la perspectiva de quien la abandona. (Williams se refiere aquí a un trabajo reciente de John McDowell.) Esto implica una forma restringida de relativismo, según la cual nuestros juicios morales sólo pueden estar de acuerdo con otros ligados a formas de vida accesibles real o imaginativamente (tal vez a través de una conexión histórica con la nuestra), de modo que puedan dar respuestas pertinentes a la pregunta de cómo debemos vivir *nosotros*. Esto también implica que la reflexión podría socavar las disposiciones prerreflexivas de las que dependen los juicios tradicionales, sin reemplazarlos con algo más objetivo.

Esto parece implicar una clase de conservadurismo ético. Pero, aunque Williams desconfía del impulso a partir de principios últimos, y basar todo en una nueva fundamentación, también cree que no puede regresar a un estado social prerreflexivo, tradicional, si es que existe tal cosa. Sin embargo, su concepción positiva del pensamiento ético me resulta oscura. El siguiente párrafo es característico:

> La crítica reflexiva debería ir básicamente en una dirección opuesta a la que promueve la teoría ética. Típicamente, la teoría busca consideraciones que son muy generales y tienen el mínimo posible de contenido distintivo, porque trata de sistematizar y quiere representar la mayor cantidad de razones posibles como aplicación de otras razones. Pero la reflexión crítica debería buscar la mayor comprensión posible sobre cualquier tema, y usar cualquier material ético que, en el contexto de la discusión reflexiva, tenga algún sentido y exija cierta lealtad. (pp. 116-17)

El estado mental hacia lo ético, lo que se supone que debemos esperar, no es el conocimiento, o aquel otro estado alternativo, propuesto frecuentemente, la decisión, sino algo distinto que Williams denomina *confianza.* La describe como un "fenómeno básicamente social", pero no nos dice mucho más que eso. Supongo que él pretende que ella debe captar esas convicciones que sobreviven a la reflexión, el análisis y la experiencia, aunque no tengan un fundamento objetivo.

Cuando personas razonables tratan de decidir qué hacer, confían en ello, lo cual es ciertamente mejor que tratar de decidir todo apelando a principios universales determinados filosóficamente. Pero esto no tiende a mostrar que nada más sea posible y, de hecho, cuando comienza a fallar la confianza o cuando se encuentra con una confianza igual pero que va en dirección contraria, nos parece natural movernos a un punto de vista más objetivo, en lugar de confiar simplemente en "lo que somos". Esta no es una tendencia filosóficamente impuesta, sino un movimiento natural del pensamiento humano, y tratamos de intentarlo, aun cuando, a menudo, no podamos pretender que resuelva nuestros problemas.

Williams reprimiría ese impulso: sostiene que es una falacia que surge del actual dominio de la racionalidad administrativa (en nuestra cultura, se supone que toda disputa tiene alguna solución defendible racional y públicamente). Williams es especialmente impaciente respecto del tipo de búsqueda de fundamentación universal que subyace al rechazo de Michael Tooley de la relevancia moral de la distinción entre aborto e infanticidio, o la acusación, cada vez más común, de "especiesismo", por nuestro trato preferencial de los humanos respecto de los animales. "Llamado, en forma más reveladora, 'humanismo', y que no es un prejuicio" dice Williams. (p. 118)

Hay un atractivo innegable en su esfuerzo de bloquear el apetito insaciable de la filosofía por las justificaciones; por exigir demasiado, se puede perder todo. Pero, si seguimos a Williams, podemos pasarnos del lado de la complacencia. Me parece que sus argumentos negativos no son adecuados para sus conclusiones radicales.

En primer lugar, la búsqueda de principios éticos objetivos podría ser representada en forma más caritativa: no tiene que ser ni tan comprehensiva ni tan reduccionista como supone Williams. El define una *teoría ética* como "un enfoque teórico acerca de lo que son el pensamiento y la práctica éticos que, o bien implica un test general para la correción de creencias y principios éticos básicos, o bien implica que no puede existir tal test" (p. 72). Pero los principios teóricos pueden ser universales sin ser totalitarios; pueden resolver algunos conflictos de valores, sin resolverlos a todos. Pueden ser regulatorios, dejando lugar a una buena dosis de pluralismo y variaciones éticas determinadas culturalmente, dentro del marco que ellos definen. La idea de que la teoría ética busca siempre la mejor forma de vida para todos los seres humanos o racionales, es simplemente un fantasma. En cierto punto, Williams sugiere que una teoría contractualista puede desempeñar un papel moral más limitado que es necesario para justificar el liberalismo, pero cree que el deseo de acuerdo razonable del cual depende puede ser lo suficientemente fuerte como para resistir la presión de valores particulares solamente en circunstancias bastante especiales, de modo que no sirve como una teoría general. Sin embargo, el pesimismo sobre las pretensiones de la teoría ética parece menos garantizado si sus ambiciones de dominación externa no son totales.

¿Cuán fuera de nosotros podemos avanzar legítimamente en nuestro razonamiento práctico, aunque sea en la búsqueda de resultados limitados? Williams cree que el punto de vista externo es fundamentalmente un punto de vista de observación, que no es la fuente de juicios independientes o evaluativos. El le atribuye la facilidad de la transición de la percepción a la ciencia al hecho de que los juicios perceptivos ordinarios, pese a su subjetividad, pretenden referirse al mundo exterior y se prestan naturalmente a la crítica y sustitución o complementación desde puntos ventajosos progresivamente más objetivos hacia el mismo mundo. Su postura es que intentar esto en el dominio práctico es un salto al vacío.

Williams encuentra esto intuitivamente obvio, pero sus argumentos parece que no pueden convencer a nadie que no piense así. La objetividad aquí tendría que ser sui géneris y no como la objetividad científica, pero esto no nos impide contar una historia sobre la razón práctica que sea, al menos, estructuralmente análoga. Nuestros primeros juicios sobre qué hacer y cómo vivir no son, en esta interpretación, meramente subjetivos, sino que son objetivos implícita o, quizás, incipientemente, en su intención. Así lo sería si hubiera, desde el comienzo, una concepción externa de nosotros mismos, que es, después de todo, uno de los rasgos esenciales de la humanidad racional, en los juicios prácticos: de modo que un juicio sobre lo que debo hacer es, por lo menos implícitamente, un juicio sobre lo que la persona que yo soy debería hacer. Esto es totalmente compatible con el reconocimiento de que el *yo* que reflexiona es también el *yo* que desea y actúa. La deliberación práctica puede ser simultáneamente de primera y tercera persona.

La visión externa no tiene que ser formulada en forma demasiado clara o consciente, a fin de dar lugar a una crítica y justificación objetiva más sofisticada. Después de todo, la nueva objetividad que hace que los juicios perceptivos continúen los resultados de la física de partículas, no anticipa las notables formas de objetividad con la que se la va a reemplazar. Pero esta analogía estructural es inútil, salvo que la noción de objetividad práctica pueda tener algún contenido. ¿Qué método general de progreso objetivo corresponde, en el dominio práctico, al método científico de explicar las apariencias como perspectivas de una realidad independiente?

La opción aristotélica es inaceptable, no sólo fácticamente, sino moralmente. Aun si reconociéramos desde afuera una naturaleza humana universal y un bien humano correlativo, esto no probaría que cada persona pueda perseguir ese bien para sí misma, o para la humanidad como un todo, excluyendo a todo el resto. La teleología aristotélica, si fuera accesible, no sería más irresistible que la sociobiología: podemos tener razones para trascender la naturaleza humana en ciertos aspectos.

Esto nos lleva a la otra opción: la objetividad kantiana. Lo que aquí corresponde al esfuerzo científico de situar nuestras perspectivas perceptuales en el contexto de una concepción absoluta de la realidad construida en principios explicativos es el

esfuerzo ético por ubicar nuestros juicios personales preteóricos sobre qué hacer en el contexto de una concepción general de lo que la gente debería hacer. Estos juicios preteóricos son considerados no como resultados de nuestra interacción causal con la realidad ética, sino como ejemplos primitivos del razonamiento evaluativo o práctico, y esperamos que las versiones, progresivamente más objetivas, generadas por la teoría ética sean ejemplos mejorados, manifestando nuestra capacidad de asumir un punto de vista universal, distinto del de cualquier persona particular.

La creencia en la objetividad implica evidentemente un compromiso con alguna clase de realismo; no realismo sobre un mundo de valores con el cual interactuamos causalmente, sino realismo sobre la existencia de razones o valores que podemos descubrir mediante ciertos procesos de pensamiento. El ejemplo de la matemática es marginalmente relevante en esta relación, no porque la matemática sea algo parecido a la ética —partes de ella gobiernan el mundo físico, después de todo— sino porque es un ejemplo de la exploración de un ámbito de verdad objetiva que no procede por medio de interacción causal con el mundo, y una construcción de una teoría causal de las apariencias preteóricas. El conocimiento matemático se explica con nuestra capacidad de seguir argumentos matemáticos, cosa que hacemos en nuestra cabeza.

En la ética también, si el realismo es correcto, la verdad de ciertos principios y la existencia de ciertas razones explicará nuestra capacidad de conocerlos por medio de nuestra capacidad de seguir las formas de pensamiento y argumento que revelan su corrección. Comenzamos con un acceso limitado a la verdad, desde un punto de vista limitado y personal, y en el curso de nuestras exploraciones nos podemos equivocar o perder. Pero podemos hacer descubrimientos, y estos descubrimientos alterarán nuestros motivos y expandirán nuestras perspectivas como agentes. La capacidad de apreciar y darle peso a las vidas de otras personas, quienquiera fuesen, es una forma de comprensión. Decir que soy yo, finalmente, el que debe actuar, no hace sino desacreditar la pretensión de tales motivos objetivos a la autoridad moral.

Ningún argumento filosófico abstracto puede negar esto: que esta es mi propia convicción sobre la ética y los límites de la filosofía. La prueba de escepticismo debe provenir de los argu-

mentos éticos mismos. Si producen ejemplos convincentes de un conocimiento extendido sobre cómo debemos vivir y tratarnos mutuamente, el conocimiento adquirido en parte por trascender nuestros puntos de partida personales y sociales, entonces la realidad de los resultados no estará desacreditada por su falta de semejanza con otros tipos de conocimiento o verdad, o por el hecho de que los individuos deben estar motivados por ellos.

Las mejores razones a favor del escepticismo en lo tocante a teoría ética, son la debilidad y controvertibilidad de sus resultados y la falta de acuerdo sobre los métodos. Hay, naturalmente, desacuerdos e incertidumbre en la ciencia y la matemática, pero también hay un enorme cuerpo de descubrimientos incontestables, que hacen que el rechazo al realismo en esos ámbitos no sea más que una fantasía filosófica. La ética es diferente, y el intento de darle su definición prematura puede llevar a una moralización no comprometida. Sin embargo, nada en el estado primitivo actual de la ética requiere que abandonemos la esperanza de la posibilidad en un eventual progreso y de la validación retrospectiva del progreso, aparentemente modesto, que algunos de nosotros creemos que se ha alcanzado. Este optimismo se combina mejor con la visión de que la teoría ética, si hay tal cosa, se encuentra en la infancia.

Por esta razón, los teóricos impenitentes pueden estar de acuerdo con Williams en algún aspecto, sobre la práctica del pensamiento ético. No deberíamos rechazar convicciones éticas que no tengan fundamentos teóricos: pueden ser aproximaciones a la verdad (*insights*), o parte de ella. De todos modos, es posible que sean lo único que tenemos. El tiene razón en que teorías demasiado abarcadoras no son sustitutos, en el actual estado de cosas, de un método de pensamiento que utilice las visiones (*insights*) y formas de crítica con las que podemos contàr, aunque el resultado sea algo anárquico. Pero esto no significa que la teoría sea imposible, aunque no podamos hacer mucho cuando tiran la puerta abajo.

18

Schelling: El precio de la vida*

Este ensayo fue originalmente una reseña de Choice and Consequence *(Harvard University Press, 1984), de Thomas Schelling. Aunque es más bien corto, lo incluyo en este volumen, debido al interés de Schelling por los monopolios ilegales, el valor de la vida y las estrategias intrapersonales de conflicto.*

Un tema central de la economía es la motivación humana en el conjunto: cómo las decisiones, racionales e irracionales de millones de personas, se combinan para producir resultados a gran escala. Descansa en una comprensión del modo como cada decisión individual está afectada por la percepción de las motivaciones de los demás, cuyos intereses pueden entrar en conflicto o superponerse con las propias motivaciones. La teoría de juegos estudia los principios racionales de tal interacción.

Thomas Schelling, un economista sutil e imaginativo, se ha destacado por haber extendido a nuevos dominios el estilo económico del pensamiento sistemático y relativamente preciso sobre la interacción humana. Su famoso libro, *The Strategy of Conflict*, publicado en 1960, trataba sobre todo la guerra, la disuasión y las relaciones internacionales. Mientras que cuatro de los quince capítulos de esta nueva recopilación de ensayos tratan acerca de la bomba y temas afines, la mayoría de ellos se refieren a otros asuntos: el valor de la vida y la muerte, el crimen organizado, las estrategias de conflicto de una persona consigo misma, y en menor proporción, la ética.

* Reimpreso con autorización de *The New Republic*, 27 de agosto de 1984.

Lo que hace que Schelling sea tan agradable de leer, aparte de su estilo sencillo, claro, irónico y coloquial, es la sagacidad y agudeza humana con que analiza problemas y fenómenos complejos, dando lugar a una comprensión sistemática de ellos. Algunas de sus observaciones tienen el encanto asociado a la obra de Erving Goffman. Por ejemplo, alguna vez nos hemos preguntado qué conexión tenía, hace unos años, el crimen organizado con el negocio de las lavanderías. Schelling propone una explicación: se trata de un camino particularmente eficaz para la extorsión, algo más refinado y mejor tanto para el criminal como para la víctima que el tradicional y tosco esquema "protector". Las típicas víctimas son los restaurantes, que pueden ser fácilmente atacados y son frecuentemente propiedad de un solo individuo. Forzarlos a aceptar un sistema de lavandería sobrevaluado y monopólico es mejor que exigirles dinero, porque (a) la víctima puede rendir la transacción en su contabilidad sin revelar los agentes del beneficio o crédito a los que le pagan, (b) no sentirá vergüenza o pérdida de autoestima, porque puede decirse a sí mismo, o a su esposa o empleados, que es meramente una víctima de los precios excesivos de la lavandería, y (c) el pago se puede deducir de los impuestos, por lo que el extorsionista puede exigir más, con menor costo por parte de la víctima.

Muchas de las víctimas directas del crimen organizado, sin embargo, son pequeños empresarios ilegales —pasadores de apuestas, prostitutas— que les pasan el coste de su pago a sus clientes. No se pueden quejar a la policía, y a veces los mismos policías están empleados directamente por el sindicato para hacer cumplir el pago, por medio de una aplicación selectiva de la ley. Schelling sugiere que el dinero real que sostiene una gran organización criminal y le hace buscar nuevas oportunidades probablemente provenga de un monopolio de mercado negro, creado por la prohibición legal de algo que tiene gran demanda. La prohibición de la novocaína, probablemente, causaría este fenómeno pero, de hecho, el alcohol parece haber impulsado al crimen organizado a un crecimiento económico, en los Estados Unidos, durante su prohibición, y ahora los narcóticos proveen la infraestructura. Schelling se pregunta "si el objetivo de algo que reduce el consumo de narcóticos, el juego, la prostitución,... o cualquier otra cosa, forzado por la ley al mercado negro, es o no mayor que los costes sociales de la crea-

ción de una industria criminal". Mi impresión es que él cree que no, pero aquí y en otros lugares, prefiere analizar los hechos que se refieren a la pregunta, en vez de responderla directamente. Si, como él sugiere, la mayor parte del daño producido por narcóticos resulta porque son ilegales y excesivamente sobrevaluados, deberían, probablemente, ser legalizados.

Una pregunta especialmente fascinante para la que Schelling está preparado para darle una respuesta tentativa es: ¿Cuánto deberíamos pedirle a la gente que gaste, o que gaste a cuenta de los beneficios por el sistema de impuestos, en políticas que reduzcan el riesgo de muerte? Esta pregunta se relaciona con la actual controversia sobre la obligatoriedad de bolsas de aire en automóviles, pero también se aplica a la salud, seguridad en los aeropuertos, salvavidas, alarmas de humo y otras medidas que salvan la vida, cuyos costes deben ser solventados por alguien, sea un consumidor o un contribuyente. Schelling sugiere que consideremos estos gastos desde el punto de vista de su valor hacia las partes, cuyo riesgo se ve disminuido y, entonces, pregunta si, desde ese punto de vista, debería haber algún modo mejor para gastar el dinero.

Esto requiere una respuesta a la pregunta de cuánto valor tiene una vida humana para quien la posee, lo que es analizado en el importante ensayo "The Life You Save May Be Your Own". Lo primero que hay que notar es que la vida de la gente pobre vale menos para ellos en términos monetarios que la vida de la gente rica, como medida de lo que estarían dispuestos a gastar para reducir el riesgo de muerte, por un valor dado. No se debe, naturalmente, a que la gente pobre valore menos la vida—todos tenemos sólo una— sino a que valora más el dinero, y tiene menos para satisfacer sus necesidades más urgentes.

Basándose en una encuesta informal sobre lo que pagarían profesionales bien retribuidos, por pequeñas reducciones incrementales en la probabilidad de muerte, Schelling concluye que ellos valoran su vida entre diez y cien veces su ingreso anual. "En números crudos, esto querría decir que un Boeing 707 lleno de hombres de negocios podría valuar esas vidas de modo que la prevención de un choque fatal de un avión completo (todavía no identificado) valdría doscientos cincuenta millones de dólares." (Nótese que, con un razonamiento paralelo, resultaría que un avión cargado de trabajadores rurales en migración valdría menos de diez millones.)

Schelling cree que el Estado no debería, en general, forzar a la gente a gastar en seguridad más de lo que ellos creen que vale: si ellos prefiriesen gastar esa suma en comida o alquiler, allá ellos. Pero, ¿qué ocurre si las medidas de seguridad se pagan con fondos públicos, o pueden asegurarse por un costo bajo sólo si se exige que todos las tengan? Schelling insiste en preguntar si valdría más, para los beneficiarios, un uso alternativo de esos fondos, de modo que la variación entre rico y pobre en el valor de vida monetario se vuelve aguda. ¿Acaso una cantidad de dinero bien empleada en la seguridad de vuelo de ejecutivos es derrochada si se la gasta en la seguridad o salud de obreros textiles o trabajadores rurales? Al determinar el valor de la vida con el fin de evaluar medidas de seguridad obligatoria o pública, ¿qué evaluaciones se deben utilizar? ¿Deberíamos tomar un promedio, o debería el Estado gastar más en la seguridad de los ricos que en la de los pobres?

Esto no es simplemente una cuestión de justicia distributiva, porque la podemos plantear aun si estamos comprometidos a exigirles impuestos mayores a los ricos en beneficio de los pobres. La cuestión es si se debe gastar dinero en algo que el pobre valora más que un determinado aumento de seguridad. El instinto de Schelling es, si no podemos darles el dinero y dejarles que decidan qué hacer con él, tratar de simular los resultados de un proceso de este tipo en los gastos públicos. Pero, si esto ocurre y seguimos viviendo en una sociedad desigual en la que los más beneficiados pagan impuestos no sólo para beneficio de los pobres, sino también para el suyo propio, podemos encontrarnos con un estándar doble para la seguridad pública.

Schelling reconoce que esto puede ser políticamente desagradable, pero su simpatía hacia la imposición de igualdad en bienes específicos parece limitada. Después de aludir al destino de las diferentes clases en el naufragio del *Titanic*, Schelling hace este comentario alarmante:

> No lo toleramos más. A quienes quieren arriesgar su vida en el mar y no pueden pagar un barco seguro no se les debería negar la oportunidad de confiar en un barco barato, sin salvavidas; pero si alguna gente no puede pagar el precio del pasaje con salvavidas, y otros sí pueden, no deberían viajar en el mismo barco.

La pregunta de si, incluso en una sociedad muy desigual como la nuestra, determinados bienes deben ser accesibles en

igual medida para todos —aun con un coste considerable de eficiencia y en preferencia de beneficios que el pobre podría valorar más— es una cuestión moral importante. Los ejemplos de Schelling la plantean vívidamente, pero su incomodidad con respecto al pensamiento moral le impide dirigirse a ella. El parece creer que, para estos propósitos, estamos todos en el mismo bote sólo si estamos, literalmente, todos en el mismo bote. Pero algunos pueden sentir que la concesión limitada a la igualdad, por su propio beneficio, que Schelling admite en el caso del *Titanic*, podría extenderse en ciertos aspectos a la sociedad como un todo, en relación con la provisión del servicio médico básico, por ejemplo.

Dos de los ensayos se refieren a situaciones en las que un individuo puede ser visualizado como si poseyera más de un yo, simultáneamente o en secuencia, cuyas relaciones mutuas pueden copiar estrategias interpersonales: el alcohólico que quiere dejar de beber, el fumador que quiere dejar el tabaco, el corredor de maratón que quiere seguir pese a que está exhausto, el moroso crónico, el desesperado que hace dietas. Las descripciones de Schelling sobre la manera en que podemos engañar, forzar y manipularnos a nosotros mismos como si fuéramos otra persona son, a menudo, agudas y divertidas. Sin embargo, sus sugerencias sobre cómo la ley puede ayudarnos a controlarnos a nosotros mismos no son muy aceptables.

Especula que podríamos llegar a reconocer promesas legalmente obligatorias para no fumar, por ejemplo, que una persona no podría dejar de cumplir. Pero, en contratos interpersonales ordinarios, en una etapa posterior, una de las partes puede liberar a la otra del acuerdo que ambas hicieron anteriormente. Si, en algún momento, somos los únicos representantes de nuestros yoes anteriores, para estos propósitos, ¿cómo podríamos negarnos la autoridad de anular nuestros yoes pasados en el caso intrapersonal? O bien soy la misma persona, o no lo soy: si lo soy, entonces la decisión es mía en cada caso. Si no lo soy, mi yo pasado no debería poder imponerme una ley para hacer algo que él quería, pero con la que yo nunca estuve de acuerdo.

El análisis de Schelling en este punto es sorprendentemente débil y carente de realismo. Creo que esto se debe a que no pensó en forma suficientemente profunda sobre el concepto de identidad personal y el papel básico que desempeña en la moral. En la mayoría de los razonamientos morales, la unidad fundamental, a los fines del beneficio o responsabilidad, es la perso-

na, extendida temporalmente, representada en cada punto en su vida por su etapa presente, con algunas excepciones en los casos de locura e inmadurez. La primacía de esta unidad ha sido atacada de manera notable por Derek Parfit, en su interesante libro *Reasons and Persons,*[1] pero requiere un enfoque más profundo que el que Schelling puede ofrecer. No son cuestiones empíricas de la clase que a él le gustan.

Los ensayos sobre estrategia y armas nucleares están precedidos por una corta introducción a la teoría de juegos, utilizada más adelante en el análisis de temas acerca de los riesgos de construir armas como "chips de regateo", y las diversas formas en que un conflicto mutuo fatal no tiene que ser un juego de suma cero: hay suficientes elecciones posibles, en las cuales ambas partes pueden perder o ganar, y Schelling trata de dar una guía analítica para distinguir el segundo caso del primero. Hay un capítulo informativo y optimista sobre si grupos terroristas pequeños pueden tener la bomba, aunque Schelling advierte que los diferentes usos de armas nucleares por parte de los gobiernos son terroristas, en todos los casos. Para que no se piense que esto tiene implicaciones morales, agrega: "No implico nada derogatorio o degradante sobre fuerzas nucleares estratégicas, al subrayar la tradicional expectativa de que su uso primario es la disuasión o intimidación, y por ello, la influencia sobre la conducta, a través de la amenaza de un enorme daño civil".

Pero, a caballo regalado no hay que mirarle los dientes. El poder analítico de Schelling, su sutileza, su lucidez y su incomodidad respecto de la ética le dan a este trabajo un sabor único. En este libro, Schelling dice cosas iluminadoras sobre el control de la renta, racionamiento de gasolina, deducciones de impuestos, eutanasia, proliferación nuclear, si el padre y la madre deben viajar en dos aviones diferentes y muchos otros temas. Independientemente de las simpatías políticas de cada uno, se trata de un medio inestimable para plantear y pensar cuestiones sociales difíciles.

Notas

1. Oxford University Press, 1984.

19

Schelling: Identidad personal y autodominio

Este ensayo fue originalmente una respuesta, no publicada anteriormente, a la Tanner Lecture, dada por Thomas Schelling en la Universidad de Michigan en marzo de 1982. La conferencia fue publicada con el título "Ethics, Law, and the Exercise of Self-Command" en Sterling M. McMurrin, comp., The Tanner Lectures on Human Values IV *(University of Utah Press, 1983) y fue reeditada en su libro* Choice and Consequence *(Harvard University Press, 1984). Estos comentarios entran más en detalle en uno de los temas que traté en mi reseña.*

La enorme importancia adjudicada a la identidad personal en el derecho y la ética puede parecer incomprensible y, tal vez, injustificada, cuando observamos cuán complejos y fragmentarios somos los seres humanos y nuestras vidas. El profesor Schelling presenta algunas razones muy interesantes para reflexionar, al exponer el caso a favor de la posición de que el conflicto motivacional intrapersonal es como una lucha entre dos yoes, que puede dar lugar a consideraciones de justicia distributiva entre ellos y también puede hacer que sea adecuado que otras partes, incluso el Estado, tomen partido en sus disputas o los ayuden a hacer valer sus mutuas pretensiones. Pero creo que será aun más difícil de lo que él sugiere reducir significativamente la importancia de la identidad personal del individuo, que comprende estos diferentes "yoes". Aun si se trata de una ficción legal o, incluso, moral, la idea de la soberanía de la persona en su totalidad tiene muchas razones a su favor.

Hay diferentes modos de asignarle importancia moral a la identidad personal: independientemente de lo que sea que haga que todos los aspectos y etapas de la vida de una persona sean *suyos*; pero creo que ellos se pueden reunir en el principio que dice que, con ciertas excepciones bastante bien definidas, cualquier parte puede representar al todo. Por partes, quiero decir o bien etapas temporales o distintos aspectos del sistema motivacional o de la personalidad, aunque en este análisis me concentraré en las etapas temporales.

Una representación de esta clase puede ser activa o pasiva. La representación activa requiere que una etapa de una persona debe poder representar al todo, hacer promesas que comprometan la acción futura, correr riesgos, cometer crímenes, o comportarse noblemente en formas que acarreen castigos o beneficios a los futuros yoes. Las condiciones de la representación activa son moderadamente estrictas. No les permitimos a los niños, locos o temporariamente trastornados que tengan responsabilidad sobre los indefinidos cursos futuros de su vida, aunque quieran tenerla, y no los castigamos con el mismo rigor por sus ofensas en etapas posteriores.

La representación pasiva significa que cualquier cosa buena o mala que le ocurra en una etapa a una persona es juzgada como afectando al todo. Esto quiere decir que se permiten compensaciones y negociaciones en vidas moralmente dudosas, cuando se imponen a lo largo de esas vidas. Estamos mucho más dispuestos a cauterizar el muñón de Ahab a fin de salvarle la vida, que salvarle la vida a otra persona.[1]

No voy a discutir este tema. La relación entre distribución de beneficios y compensación en esas vidas y a lo largo de ellas, y la influencia que esta relación muestra que la identidad personal ejerce sobre problemas de justicia distributiva, han sido muy analizados recientemente, desde que Derek Parfit abrió el tema. Quisiera decir lo siguiente: si seguimos considerando que las personas completas son la unidad moral fundamental, como creo que ocurre, habrá lugar para pautas de justicia distributiva a lo largo de las vidas que no se apliquen en las vidas, tal como los principios igualitarios. Pero si se rebaja la identidad personal, no está claro si esto resultaría en menor compensación dentro de vidas y mayor compensación a lo largo de las vidas, es decir, si las restricciones de negociación que ahora son adecuadas para el grupo deberían aplicarse a dife-

rentes etapas de una persona, o si negociaciones que ahora son permisibles dentro de una vida deberían aplicarse al grupo.

Quisiera concentrarme en la representación activa, porque allí nos encontramos con los problemas fascinantes planteados por Schelling. El nos pide que consideremos qué notable es que uno no pueda hacerse a sí mismo una promesa legalmente vinculante, y especula sobre cómo cambiar esto para aumentar las posibilidades de autodominio. Empezaré considerando la imposibilidad de hacer promesas vinculantes, ni a sí mismo, ni a nadie. Tal como dice Schelling, no es autoevidente que el derecho a firmar un contrato es un derecho que debería ser ampliamente reconocido. El dice que esta es una cuestión empírica, pero si tomamos seriamente la imagen que sugiere de la persona como una multiplicidad de yoes que se suceden y alternan entre sí, entonces resulta un problema moral, además de tener consideraciones empíricas.

Si firmo un contrato o hago una promesa, estoy comprometiendo a mis futuros yoes a hacer algo. Supóngase que tomo un préstamo de mucho dinero, que debo devolver en los próximos dos años, bajando así el estándar de vida de esas futuras etapas. No importa si quiero el dinero para unas vacaciones placenteras o para mi obra de caridad favorita. Si yo trabajara y ahorrara ese dinero durante dos años y lo gastara al final de ese período, no habría problema: sería un caso de altruismo voluntario hacia mi yo futuro o hacia la obra de caridad. Pero, al invertir el orden, resulta más bien un caso de servidumbre involuntaria si pensamos en términos de yoes sucesivos. Naturalmente, el yo posterior que tiene menos dinero puede estar igualmente contento, porque está agradecido por el recuerdo de sus vacaciones o porque todavía cree que la caridad es una causa valiosa. Pero no importa si esto es así. Aun si odié mis vacaciones, o estoy desilusionado por la caridad, mi yo posterior puede estar obligado a pagar el préstamo que el yo anterior ha recibido.

No toleraríamos un sistema que le permitiera a una persona hacer una promesa exigible que a su vez fuera vinculante con otra promesa: sería una forma de esclavitud. ¿Por qué lo toleramos aquí? Lo toleramos porque la unidad importante es la persona total, y consideramos que el yo que promete y también el yo siguiente que actúa son los representantes legítimos del todo. De modo que, si el segundo se lamenta por el peso del

pago de las deudas, la réplica es (dadas ciertas condiciones de competencia ya mencionadas) que *él* se metió en eso, *él* recibió el dinero, *él* se fue de vacaciones, y si no las disfrutó, fue un riesgo que *él* decidió correr. *El* aquí no es la etapa posterior o anterior, sino la persona total, de la cual ambos son los representantes.

Hay algo artificial en todo esto, y tal vez debería haber más límites en el derecho a hacer contratos de los que se establecen generalmente. Las leyes de divorcio liberalizadas van en esta dirección: ahora es imposible, en la mayoría de los registros, sellar un matrimonio que sea irrevocable para toda la vida de los cónyuges. Es posible que esto refleje la convicción de que la vida es demasiado complicada como para permitirles a los hombres y mujeres a los veinte años una autoridad tan absoluta sobre la forma personal de su vida en sus próximos cincuenta años, aunque ellos quisieran ejercerla. Pero, en cuestiones menos personales, yo esperaría que los contratos superaran las dudas sobre identidad personal, sobre todo porque el yo posterior, aunque se sienta explotado por el yo anterior, cuyas deudas tiene que pagar, puede causarle deudas similares a *sus* sucesores y puede mantener su sistema de créditos cumpliendo las promesas del pasado, así como un gobierno nuevo cumple los compromisos de un régimen anterior odiado.

¿Por qué es diferente una promesa a mí mismo? Si puedo comprometer a mi futuro yo a hacer pagos bancarios, ¿por qué no puedo hacer yo una promesa legalmente vinculante a no fumar, beber o comer patas de cerdo asadas? ¿Por qué puede mi yo posterior liberarme de promesas de este tipo?

La respuesta simple es que no hay ninguna segunda parte involucrada. Pero si nos tomamos en serio la idea de los múltiples yoes, no se ve tan claro por qué esta respuesta es aceptable. ¿Por qué no desempeña uno de los yoes, el yo al que se le hace la promesa, el papel de la parte cuyos intereses se garantizan al hacerla cumplir?

Para ver la razón, debemos volver al caso de la promesa al otro, vista en términos de múltiples yoes. Tal promesa involucra tres partes: el que hace la promesa, al que se le hace la promesa (destinatario) y el que la cumple (ejecutante). El que promete y el ejecutante son etapas de la misma persona, y el primero puede obligar al segundo. Sólo el tercero, el destinatario, puede eximir al segundo.

¿Cuál sería la estructura correspondiente en el caso de la promesa a sí mismo? También involucra a tres partes —el que promete, aquel a quien se le hace la promesa (destinatario), y el que la cumple (ejecutante)—, pero serían todas etapas de una persona: el que hace la promesa y el destinatario serían simultáneos o, quizás, idénticos, y el ejecutante sería posterior.

Si esto funcionara como una promesa a otra persona, sólo el destinatario podría eximir al ejecutante de su obligación. Si identificamos al destinatario con una etapa anterior, entonces él no está más en el momento del cumplimiento, especialmente si el ejecutante no tiene ningún deseo de cumplir con el contrato. Sin embargo, esto no es, por sí mismo, un obstáculo para hacer valer la promesa, ya que en general la muerte de un destinatario no cancela la obligación de cumplir, tal como sabemos por las condiciones estrictas, que van más allá de la vida de un donante que hace donaciones a museos. El interés de un destinatario muerto puede garantizarse por medio de un apoderado.

¿Por qué no puedo hacer lo mismo, y prometerme a mí mismo no beber nunca más? El que quería que yo dejara de beber puede no existir más, pero, ¿por qué no hay un cuerpo de apoderados, o el Estado, que se encarguen de hacer que cumpla mi compromiso con él? La respuesta es que nadie puede ser un apoderado adecuado para los intereses de mi yo pasado, como mi yo presente. Si mi yo pasado tiene la autoridad, en virtud de alguna conexión personal, para comprometer a mi yo presente a realizar tareas onerosas, entonces mi yo presente, en virtud de la misma conexión personal, debería tener la autoridad de revisar los compromisos hechos por mi yo anterior, y eximir al ejecutante de su obligación parecería deseable, aunque el ejecutante sea yo u otra persona. Las dos direcciones temporales de la autoridad deben estar juntas: cualquier otra cosa sería sumamente injusta.

Los yoes posteriores están en desventaja con respecto a los anteriores porque estos pueden comprometerlos, sin que ellos puedan comprometerlos a su vez. Por el otro lado, están compensados de algún modo, al poder eximir a otras partes de compromisos tomados con yoes anteriores, en tanto que los yoes anteriores no pueden olvidar promesas hechas a los yoes posteriores, por más que quieran. Ni tampoco puede un adoles-

cente tener la indulgencia que su yo posterior tendrá con sus hijos.

Pero describir todo esto en términos de las relaciones entre yoes componentes es engañoso, porque oscurece la importancia de la identidad personal en la cuestión. La única razón por la que un yo anterior puede comprometer a un yo posterior, o un yo posterior puede perdonar a uno anterior, es que ambos son representantes de la persona completa, de todos sus intereses en el transcurso del tiempo. Ellos pueden no ser totalmente competentes, o ser representantes imparciales, pero, salvo en ciertas condiciones de extrema incompetencia, sea temporaria o permanente, que justifique que otro asuma como apoderado, ellos son considerados razonablemente como los custodios legítimos de *su vida*. Soy *yo*, y no mi yo pasado, el beneficiario de la promesa que me hice en el pasado, y soy yo, actuando en el presente, el que puede eximir a alguien de esa promesa. Similarmente, soy yo, y no mi yo presente, el que puede hacerse una promesa para cumplir en el futuro, y soy yo actuando en el futuro el que debe cumplirla. Sin un fuerte elemento de identidad personal, por más problemático que sea, nada de esto tiene sentido.

Después de la infancia, ningún momento tiene más atribuciones que otro para representar al yo completo y, trastornos aparte, seguimos un crudo principio del derecho de posesión. Porque, tarde o temprano, todo momento tiene que ser ahora: esto da lugar a una cierta igualdad, en especial porque no permitimos que la gente se venda como esclavos.

La única alternativa aceptable a este tipo de autonomía sería un aumento de paternalismo por parte del derecho o de la sociedad, la intrusión más frecuente de juicios objetivos respecto de si alguien representa adecuadamente los intereses de su vida como un todo. No creo que sea una manera creíble de lograrlo darle al presente más autoridad para aplicar la ley, para hacer valer su voluntad en el futuro. Schelling señala que el "yo equivocado" podría llegar a tener control legal. Creo que la única solución a este problema sería imponer patrones externos, pero no soy optimista respecto de la posibilidad de formular alguno que no destruya más libertad de la que pueda crear.

El derecho a hacer los compromisos consigo mismo legalmente vinculantes de cierta clase socialmente aceptados —a hacer gimnasia, a no fumar, a no beber— sería una forma debilitada

de paternalismo. Pero no estoy seguro de que los argumentos a favor sean mucho más fuertes que los argumentos a favor de un paternalismo franco en estas áreas. Hay elecciones que no sólo es mejor dejárselas a los propios individuos, sino que también es mejor que queden abiertos a revisión o renovación en el curso de la vida.

Tal vez yo esté solamente revelando mis dudas sobre que las dietas realmente valgan la pena, ya que no objeto al paternalismo en seguridad social o en el uso de los cascos para motociclistas. Pero creo que quienes más probablemente se beneficien de instituciones de compromisos opcionales son los que más protección necesitan de las ambiciones rigurosas del yo paternal.

Notas

1. Se trata del ejemplo de Schelling, tomado de la película *Moby Dick.* Hay que sujetarlo a Ahab a la fuerza para cauterizar el muñón de su pierna amputada. (*Choice and Consequence,* p. 83.)

20

Dworkin: Interpretación y derecho*

Este ensayo fue originalmente una reseña de Law's Empire, *de Ronald Dworkin (Harvard University Press, 1986).*

Esta gran obra teórica aparece en un contexto político definido. En los Estados Unidos las teorías jurídicas son políticamente controvertidas. El público está muy consciente de que el modo como se interpreta la ley, especialmente en las cortes federales de apelación y en la Corte Suprema, tiene y tendrá importantes consecuencias en su vida y libertades. La controversia surge no sólo sobre temas específicos, como el rezo en las escuelas, censura, aborto, discriminación inversa y los derechos de los criminales acusados, sino también sobre los tipos de fundamento en base a los que se decidirán casos relacionados con estos temas. Muchos de los ataques verbales que surgen a raíz de cada decisión importante de la Corte o de cada nominación del ala derecha para un escaño en el gobierno tienen un carácter específicamente filosófico: los liberales acusan a los conservadores de no reconocer los derechos individuales; los conservadores acusan a los liberales de inventar la ley en lugar de descubrirla.

Al emerger de la misma cultura legal, la jurisprudencia filosófica americana está también entretejida en la controversia política y, en el nivel teórico, en la filosofía moral y política. El

* Reimpreso con autorización de *The London Review of Books*, 18 de septiembre de 1986.

filósofo del derecho americano más distinguido y original, Ronald Dworkin, fue profesor de jurisprudencia en Oxford durante muchos años, pero su corazón y gran parte de su influencia, se encuentra en los Estados Unidos. Pese a que, a menudo, utiliza ejemplos del derecho británico, las cuestiones fundamentales que más le ocupan se refieren a la igualdad y los derechos de los individuos ante el poder del Estado, que son especialmente conspicuos en el escenario jurídico norteamericano.

Dworkin es un liberal en el sentido norteamericano, uno de los representantes intelectuales más prominentes de esa posición. Es la clase de persona que los conservadores estadounidenses tienen en mente cuando acusan a los liberales de pedirles a los jueces que ignoren lo que es la ley y que la sustituyan con sus opiniones personales sobre como debería ser: es decir, un abuso de poder y un engaño del proceso democrático. Se supone que los jueces conservadores no permitirán que sus opiniones políticas influyan su práctica judicial, sino que sostendrán firmemente lo que la ley dice.

Esta es una forma común de hipocresía: invocar la autoridad de una posición ostensiblemente neutral y de orden superior, como apoyo de las propias convicciones sustantivas y partidarias. Pero la relación entre sus teorías generales del derecho y sus posiciones específicas sobre casos controvertibles es un problema importante y difícil. A Dworkin se lo identifica, por ejemplo, con posiciones fuertemente liberales sobre la permisividad de la discriminación inversa y la no permisividad de la supresión de la pornografía. También se lo identifica con la posición de que el razonamiento moral cumple una función crucial en la interpretación jurídica. Quienes están en desacuerdo con él en los temas sustantivos pueden estar tentados de aducir que su teoría de la adjudicación es sólo una excusa para atribuirle al derecho sus propias preferencias morales y políticas.

Una de las muchas virtudes de su nuevo libro es que nos permite examinar y decidir sobre esta acusación. *El imperio de la justicia* [Barcelona, Gedisa, 1988] es una obra rica y compleja: trae a luz distinciones entre los diferentes lineamientos del pensamiento de Dworkin en derecho, política y moral, y las formas en que ellos se encuentran en relación de independencia entre sí. Hasta ahora, Dworkin ha escrito ensayos, muchos de los cuales se encuentran compilados en dos libros: *Taking*

Rights Seriously[1] y *A Matter of Principle.*[2] Pero este libro representa el primer tratamiento extenso de su teoría general del derecho, y el primer libro que escribió desde el principio como tal. Se trata de un libro ambicioso, que satisface las expectativas de una gran obra escrita por un gran pensador. Dworkin ha desarrollado un sistema de ideas fuerte y complejo, que se exponen aquí con la claridad y elegancia a la que sus lectores ya están acostumbrados.

Los jueces ejercen inevitablemente un gran poder en cualquier sociedad, en la que se resuelven discusiones importantes mediante la apelación al derecho: y tanto mayor cuando las leyes pueden ser revocadas por la apelación a una constitución escrita. Esto se debe a que el derecho debe aplicarse a casos reales, cuya gama de variaciones es interminable en muchas dimensiones, y no hay ningún código significativo, cláusula constitucional o decisión previa, por más cuidadosamente diseñados que estén, que puedan decir cómo aplicarlo en cada uno de los casos posibles. Frecuentemente, la aplicación puede ser clara, pero a veces, y a menudo en casos muy importantes, se debe interpretar el derecho para poder aplicarlo, y no es obvio cuál es la interpretación correcta. Para ser juez, no alcanza con saber leer.

El primer punto de Dworkin es que cuando los jueces están en desacuerdo sobre la decisión correcta en un caso difícil, están en desacuerdo sobre la interpretación correcta del derecho, es decir, a qué se refiere ese derecho. En tales casos, no hay un hecho simple respecto de lo que es el derecho, que cualquiera, dotado de suficiente información, pueda descubrir. Pero esto no quiere decir que no haya ningún derecho en absoluto, tal que los jueces puedan inventarlo. El derecho, como dice Dworkin, es un concepto interpretativo. Para decidir lo que el derecho requiere en un determinado caso, es necesario considerar no sólo los hechos del caso, el "lenguaje llano" de la ley, y los ejemplos de su aplicación previa, sino también el sentido del derecho y de las instituciones y prácticas a las que pertenece. Todo esto forma parte de la práctica de la adjudicación.

Al justificar una decisión en un caso controvertido, es inevitable la confianza en alguna concepción de la función de los tribunales en el orden político, su relación con el poder legislativo y ejecutivo, y los principios que garantizan el uso del poder del Estado para limitar o proteger la libertad individual.

El argumento legal presupone siempre una fundamentación jurídica, aunque sea encubierta. Y, dado que el papel del poder judicial puede justificarse sólo en términos de una concepción más amplia del orden político-legal, presupone también una moral política. Dworkin resume el sentido fundamental de una práctica jurídica del siguiente modo: "El derecho insiste en que no se debe usar o impedir la fuerza, por más que sus fines deseados sean útiles, por más que esos fines sean beneficiosos o nobles, excepto tal como lo permiten o requieren los derechos y responsabilidades individuales que emergen de decisiones políticas pasadas respecto de cuándo está justificada la fuerza colectiva".

Sin embargo, la jurisprudencia no es idéntica a la teoría política, que puede ser utópica, en tanto que la jurisprudencia no puede serlo. Una teoría política puede tratar de describir un orden político y jurídico ideal, pero una de las tareas más importantes de la jurisprudencia es explicar por qué los jueces deben aplicar y hacer valer leyes que ellos no aprueban, o contra las que hubiesen votado y que consideran rechazables. Un sistema jurídico con supremacía legislativa y una doctrina de respeto hacia el precedente ubica siempre a los jueces frente a esa responsabilidad. Más aún, surge no sólo donde el sentido del derecho es incontrovertido, sino también en los casos en que se necesita una interpretación. Y, dado que la interpretación inevitablemente implica juicios sobre la mejor manera de realizar el sentido o propósito de una ley y del sistema al que pertenece, es una tarea particularmente delicada definir de qué modo el punto de vista del juez puede combinarse con otros factores, para dar lugar a una conclusión sobre lo que esa ley es, que puede no corresponder con lo que él cree que debería ser.

La respuesta de Dworkin a esta pregunta, que es el núcleo teórico de su libro, es una concepción que denomina "El derecho como integridad". Al principio, pensé que este título pretendía solamente darle a la teoría un barniz de santidad. (Las posiciones opuestas se llaman "convencionalismo" y "pragmatismo".) Pero esta denominación tiene una gran ventaja. Consideremos cuán ambigua es la integridad como virtud. Puede adscribirse respetuosamente incluso a alguien cuyos principios rechazamos y cuyos propósitos son opuestos a los nuestros. Una persona con integridad es alguien cuya con-

ducta se deriva de sus principios, a pesar de la opinión pública, presión oficial, tentación personal, cuya conducta forma cierta clase de todo moralmente inteligible, aun si sus valores son incorrectos.

Este no es un mal modelo para la interpretación judicial: una construcción del derecho y sus propósitos de modo que haga que las decisiones emerjan de un conjunto coherente de principios, aunque esos principios no sean los nuestros. La finalidad imaginada es personificar el sistema legal, de modo que cada juicio sea concebido como una manifestación de un punto de vista coherente y penetrante, aunque complejo. Un juez, en estas decisiones, está obviamente limitado por el requisito de adaptarse al lenguaje del derecho y a las decisiones previas, y tiene que aceptar ciertas incompatibilidades entre estos antecedentes. Pero el problema difícil es decir qué significa seguir adhiriendo al principio cuando la adaptación por sí misma no determina un resultado claro, es decir, cuando puede considerarse como apta a más de una decisión posible, algunas tal vez mejores que otras. ¿Qué clase de argumento se puede utilizar para identificar la decisión correcta?

Según Dworkin, la interpretación jurídica debería tratar de descubrir la interpretación del derecho que hiciera su puesta en práctica justificable a la luz de la mejor justificación que se le puede dar del sistema total del derecho y del precedente al que pertenece, lo cual incluye la mejor justificación de la función del juez en dicho sistema. Esta tarea no es solamente autorreferencial, sino también normativa: el mejor sentido del derecho es el sentido que, relativo a las limitaciones de adaptación a lo ya dado, hace que el sistema sea el mejor posible, el mejor, tanto en adaptación como en sustancia, equilibrado de forma que sea en sí mismo objeto de interpretación evaluativa. Los argumentos sobre esta cuestión inevitablemente se referirán a las convicciones morales y políticas básicas de los jueces, no porque se pueda suponer que estas convicciones sean correctas, sino porque los temas políticos y morales forman una parte importante de lo que el juez cree al llevar a cabo la tarea de interpretación. (En este aspecto, no es diferente de la confianza que el jurado tiene que depositar en sus convicciones empíricas al evaluar la evidencia.)

Por ejemplo, después de varias decisiones divididas de la Corte Suprema, sigue siendo un tema en el derecho estadouni-

dense si la discriminación inversa compensatoria —que las instituciones públicas empleen preferentemente a negros— es inconstitucional por el mismo criterio de "igual protección de las leyes" que se usó para declarar inconstitucionales a la antigua discriminación contra los negros por parte del Estado. Una cuestión crucial es si la cláusula de igual protección de la Enmienda 14ª requiere *(a)* que la *categoría* de raza no debe usarse para asignar tratamiento diferencial, o *(b)* que el tratamiento diferencial no se base en cierto tipo de *razón*, a saber, el prejuicio racial. La discriminación racial tradicional sería descartada por cualquiera de estos principios: ni el lenguaje de la Constitución, ni el requisito de adaptación con el precedente decide entre ambos. El problema es, entonces, cuál interpretación de la cláusula de igual protección, como un obstáculo hacia la discriminación racial, permite una mejor lectura de ella en el contexto del sistema legal estadounidense. Esto es parcialmente un juicio moral, pero tiene un aspecto jurídico: el principio *(a)* reglamentaría a la discriminación inversa como inconstitucional; el principio *(b)* no lo haría. Dworkin sostiene que *(b)* es la lectura correcta, es decir, que deberíamos penetrar en la razón, más allá de la categoría, para explicar lo que está mal en la discriminación.

Dicho método parece darle al juez libertad total para insertar su política dentro del derecho, en casos controvertidos, para interpretar la ley como lo que él hubiera votado como legislador. Pero esto no es así, porque el papel de los tribunales es, en sí mismo, una cuestión de interpretación jurídica, y la mejor lectura de su papel en el sistema político en general no le garantiza una discreción total en los casos controvertidos. En otras palabras, dadas las restricciones de adaptación al derecho escrito actual y los precedentes, la mejor interpretación del sistema, que permite su mejor lectura moral y política, requiere que los tribunales tengan mucho más respeto hacia la coherencia con el derecho preexistente y periférico, que lo que se requiere del poder legislativo.

El marco general de Dworkin deja lugar para una gama de posiciones sobre el tema de restricción judicial. Deberíamos notar una distinción útil que él establece entre dos sentidos muy diferentes del contraste "liberal-conservador", que se confunden frecuentemente en la discusión pública. Un juez puede ser conservador en el sentido de que sus convicciones sobre la

adaptación son estrictas, dejando el menor lugar posible para que valores sustantivos ejerzan influencia en la elección entre interpretaciones, pero un juez que es conservador en este sentido puede ser liberal en sus convicciones políticas y morales —sobre la justicia y la libertad individual, por ejemplo— y esto habrá de influir sobre sus interpretaciones, dentro de los límites que le permitan sus convicciones estrictas sobre adaptación. Similarmente, un juez que es políticamente conservador puede ser liberal o incluso radical respecto de la adaptación; en tal caso, será un activista jurídico que se permite a sí mismo la libertad para decidir casos en una dirección conservadora, derrocando lo precedente cuando la letra de la ley no lo excluye claramente. (Esto es cierto del juez Rehnquist, por ejemplo, que será elevado a Presidente de la Corte, cuyas opiniones de la Corte Suprema están fuertemente "orientadas hacia los resultados": son fácilmente predecibles sobre la base de sus posiciones políticas personales.)

Las convicciones políticas sustantivas de Dworkin son liberales: un igualitarista económico y social, comprometido con derechos individuales fuertes en contra de la voluntad de la mayoría en el cumplimiento de la vida personal. Respecto del rigor de adaptación, es moderadamente liberal, otorgándoles peso significativo, mas no dominante, a tales factores. Pero dentro de la concepción general del derecho, en tanto integridad, como él reconoce, su posición particular podría ser atacada por ser demasiado liberal o demasiado conservadora respecto de la adaptación, o por no ser demasiado conservadora, o no suficientemente radical respecto de la sustancia moral.

Quisiera describir un área de controversia, a modo de ilustración. Alguien menos complaciente que Dworkin hacia la tan difundida protección de los derechos individuales contra el poder legislativo y ejecutivo en la interpretación constitucional estadounidense en las últimas décadas, podría argumentar lo siguiente, combinando diversos puntos. Podría decir, en primer lugar, que se muestra un respeto insuficiente hacia la intención de los autores de la Constitución. Evidentemente, no tenían en mente un derecho de privacidad de la clase que se usa para revocar las leyes que prohibían el aborto. Tampoco tenían en mente los autores de la Primera Enmienda que la publicación de pornografía sería protegida por la libertad de

prensa. El podría decir, también, que incluso si la intención de los autores y del precedente pasado no descartaban dichas interpretaciones es, en tales casos, contrario a principios fundamentales de la democracia oponerse a la voluntad de la mayoría, tal como el legislativo lo expresa, sobre la base de interpretaciones controvertidas hechas por jueces no elegidos, por medio del razonamiento moral y filosófico. El poder de revocar leyes votadas democráticamente debería ejercerse poco, y sólo sobre el fundamento textual más fuerte. Por último, él podría decir que muchos de los "derechos" que fueron reconocidos en decisiones recientes son moralmente espurios, y que un sistema que los protege es peor que uno que no los protege, aun dejando de lado cuestiones de historia constitucional.

Dworkin tiene varias respuestas respecto de la intención, pero la más importante es la siguiente: aun si dejamos de lado cuestiones tales como de quién son las intenciones que debemos considerar y cómo tenemos que descubrirlas, qué tenemos que hacer si están en conflicto entre sí, y qué hacer en el caso de que nunca se le prestó atención a la cuestión que estamos tratando de decidir ahora, queda pendiente el problema de que cualquier autor de un estatuto o disposición constitucional puede tener intenciones tanto abstractas como concretas que, en nuestra opinión, están en conflicto, y tendremos que decidir cuál de ellas debemos interpretar como dominante.

Por ejemplo, para volver a la cláusula de igual protección: cuando fue adoptada, no se la pensó para descartar escuelas públicas racialmente discriminadas. El Congreso que propuso la Enmienda 14ª en 1866 siguió manteniendo escuelas segregadas en Washington D.C. (que estaba bajo su responsabilidad directa), y siguieron siendo discriminadas hasta que la discriminación fue promulgada inconstitucional sobre la base de dicha enmienda en el caso de *Brown v. Board of Education*, ochenta y ocho años más tarde.

La decisión *Brown* puede explicarse como el descubrimiento de que la intención abstracta de los constituyentes, garantizar igual protección de las leyes, era incompatible, contrariamente a su propia creencia, con su intención concreta de permitir escuelas públicas segregadas, y una decisión de tomar la intención abstracta como la dominante. (Esto es un ejemplo también de la distinción entre concepto y concepción: el concepto de igualdad de trato puede interpretarse según las diferentes concepciones de cómo debe llevarse a cabo.)

El juicio de incompatibilidad no es lógico, sino ético-político, que utiliza ideas sobre la raza y sociedad que han sido ampliamente aceptadas a lo largo del tiempo. Dworkin cree que está bien tomar la intención abstracta, y no la concreta, como dominante, tanto en este como en otros casos de conflicto, guiándose por lo que le da el mejor sentido moral al sistema como un todo. En efecto, sostiene que su permiso para corregir las intenciones concretas con referencia a las abstractas está de acuerdo con las intenciones de los constituyentes mismos (quienes no pensaban que eran infalibles en cosas tales como lo que constituye el igual trato), y también está autorizado por una larga tradición de práctica jurídica. Si esto es correcto, los tribunales de apelación tienen una libertad considerable para usar razonamientos morales en la interpretación de expresiones abstractas como "castigo cruel e inusual", "libertad de prensa", "establecimiento de una religión" y "debido proceso ante la ley."

Pero esto nos lleva a la cuestión de si el darles a los jueces no elegidos tanto poder como para socavar la voluntad de la mayoría es antidemocrático, de un modo que deja de ser la mejor interpretación del papel de lo jurídico en el sistema legal estadounidense, ya sea desde el punto de vista de la adaptación a las tradiciones y prácticas estadounidenses, o desde el punto de vista de la moral política. Esto trae consigo un problema de teoría política muy difícil —la función que cumplen las restricciones a la regla de la mayoría, al garantizar la legitimación del gobierno democrático— y trae inevitablemente a luz el tercer punto del argumento: el escepticismo respecto de la fuerza o realidad de ciertos derechos individuales.

La posición de Dworkin en estos temas, y su comprensión de la integridad en el derecho, se deduce de una concepción general de la legitimación política: qué significa para una comunidad reclamar una lealtad especial a sus miembros, y su obediencia a sus decisiones colectivas como obligación y no mero poder. En un interesante análisis sobre la ética de la asociación, Dworkin argumenta que una verdadera comunidad de principio debe ser más que un marco de compromiso político entre intereses opuestos, y que la legitimidad requiere que lleve a cabo tanto la virtud de la justicia —tratar a los miembros de la comunidad como iguales— como lo que llama equidad en forma algo idiosincrásica: darles a los ciudadanos una influencia más o menos igual en las decisiones que los gobiernan. Ob-

viamente, estos pueden entrar en conflicto, como cuando en una votación democrática una mayoría trata de suprimir una religión minoritaria: la minoría no es tratada equitativamente, aun cuando tenga un voto como cualquier otro. El equilibrio entre estos dos elementos en un sistema legítimo, que a menudo adquiere la forma de un equilibrio entre la regla de la mayoría y los derechos individuales, puede delinearse de varias maneras. Pero, según cree Dworkin, los tribunales pueden suponer legítimamente la responsabilidad de mantener la integridad del sistema.

Algunos discutirían que la igualdad de la influencia política es, relativamente, mucho más importante para la legitimación de lo que Dworkin cree, y que muchos de los derechos protegidos por la Constitución de los Estados Unidos han sido diseñados, en realidad, para mantener a las minorías vulnerables privadas de ella. Dirían que las decisiones controvertidas sobre los requisitos de la justicia no deberían quedar en manos de legisladores electos —como ocurre en Gran Bretaña— en vez de en manos de los jueces. Esto no quiere decir que la legislación debería ignorar al argumento moral sobre cuáles derechos individuales necesitan protección respecto de las pretensiones de bienestar general o voluntad popular, sólo que la autoridad para proporcionarles a esos derechos la protección de la ley debe fundamentarse más ampliamente de lo que tiene validez en una corte de apelación no electa. (Esta afirmación es menos aceptable con respecto a la protección de minorías discretas.)

Dworkin respondería, posiblemente, que no se debe confiar en la voluntad popular o, incluso, en el legislativo, para determinar los límites de su propia autoridad. Pero se podría pensar que no se debe confiar tampoco en la Corte Suprema para determinar los límites de *su* autoridad, que es, precisamente, lo que ocurre en el sistema actual. Esto conduce, en la interpretación de Dworkin, a suponerle legítimamente a la Corte un enorme poder para revocar la legislación sobre la base de razones parcialmente históricas, aunque sean, en su mayor parte, morales. A pesar de que el carácter del fundamento moral es limitado, esta pretensión de legitimidad sigue siendo radical, y el lector puede sentir que los argumentos de Dworkin no alcanzan a reducir la base del escepticismo sobre esta división especial de trabajo entre las cortes y el legislativo.

Dworkin es cuidadoso al decir que no pertenece a la teoría política utópica y que, tal vez, un sistema muy diferente de nuestra mejor interpretación podría ser más acertado. Sin embargo, es evidente que sus convicciones filosóficas sobre los derechos individuales, y el papel de la filosofía moral para descubrirlas, cumplen una función importante para convencerlo de cuál es la mejor interpretación del derecho anglonorteamericano, como lo sería en la teoría del derecho como integridad. Otros, con diferentes convicciones sobre la moral y legitimidad política, pueden defender diferentes interpretaciones dentro del mismo marco. Esto ayuda a hacer más inteligible el carácter de los debates reales sobre jurisprudencia, y ubica certeramente las causas de desacuerdo.

En el libro aflora algo que sería mejor no tener, como un yuyo en el césped. Dworkin tiene una debilidad misteriosa por analogías entre la interpretación jurídica y literaria; más aún, compara seriamente la sentencia con escribir el capítulo siguiente en una novela secuencial, lo cual no apoya demasiado su posición de que hay una respuesta correcta en casos difíciles. Dado que sus afirmaciones generales respecto de la interpretación son mucho menos creíbles que sus afirmaciones específicas acerca del derecho, es una suerte que su caso no se apoye en paralelismos como este: "la interpretación de obras de arte y prácticas sociales, sostendré, se ocupa esencialmente del propósito, y no de la causa. Pero los propósitos en juego no son (fundamentalmente) los del autor, sino los del intérprete. Básicamente, la interpretación constructiva es cuestión de imponerle propósito a un objeto o práctica, a fin de convertirlo en el mejor ejemplo posible de la forma o género al que se considera que pertenece". Esto sugiere que tenemos la custodia de las obras de arte producidas por otros, así como de nuestras prácticas e instituciones sociales. Quienes lo encuentran inaceptable, como yo, pueden extraerlo seguramente de los argumentos principales del libro.

Dworkin argumenta también, en forma poco convincente, que la mejor interpretación de la ley común de compensación económica por daños implica un ideal de igualdad en la distribución de recursos, antes que en la eficiencia o en una mezcla de eficiencia y libertad. Aquí me parece que realmente está tratando de encontrar en el derecho sus propias convicciones, olvidando rápidamente otras alternativas. En todo caso, el ar-

gumento está muy comprimido y depende demasiado de referencias a su largo y complejo análisis sobre la igualdad en *Philosophy and Public Affairs* (1981). Las posiciones de Dworkin, muy interesantes, sobre la justicia distributiva, todavía en proceso de desarrollo, no están adecuadamente representadas aquí. Pero la mayor parte del libro es muy completa y es muy accesible, pese ser un trabajo que trata cuestiones abstractas en un nivel tan elevado. El libro está lleno de sustancia y es admirablemente apto para el papel importante e influyente que, sin duda, va a desempeñar.

Notas

1. Harvard University Press, 1977.
2. Harvard University Press, 1985.

21

MacIntyre contra la Ilustración*

Este ensayo es una reelaboración de una reseña de Whose Justice? Which Rationality? *de Alasdair MacIntyre (Duckworth / University of Notre Dame Press, 1988). Nunca estoy seguro de si MacIntyre está expresando sus creencias reales, o si sólo está siendo provocativo, pero la agresividad de esta reseña se debe, en parte, a mi propia convicción de que él tiene la inteligencia suficiente como para darse cuenta de lo que está mal en sus propios argumentos y que, simplemente, no se esforzó más en ello, porque lo haría atrasarse en su trabajo.*

Cada empresa humana tiene su historia, y los que participan en ella se encuentran con la pregunta de cómo relacionar su participación actual con esa historia, incluyendo su eventual comprensión del presente. Una autoconciencia histórica excesiva puede ser paralizante pero, en alguna medida, es inevitable, y sus problemas son especialmente agudos en el campo de la ética. Al reflexionar sobre lo que es correcto e incorrecto, cualquier persona reflexiva sabe muy bien que su concepción de esos temas, los argumentos relevantes y el lugar de la carga de la prueba está fuertemente influida por el clima moral que la rodea. Sin embargo, no podemos simplemente aceptar esto sin ningún cuestionamiento, así como aceptamos la influencia de la cultura lingüística para el uso, gramática y estilo que determinan la elección de nuestras palabras. Queremos escribir y hablar correctamente, dentro del marco de la lengua de nuestro lugar y época, pero no hay ningún lenguaje *correc-*

* Reimpreso con autorización de *The Times Literary Supplement*, 8-14 de julio de 1988.

to. En cambio, actuar correctamente según los patrones de alguna moral no es suficiente: tendríamos que ser muy haraganes moralmente para no preocuparnos por la posibilidad de que la moral dominante en nuestra cultura contenga algo fundamentalmente equivocado. Pero, ¿cómo debemos entender la idea no relativista de verdad en su aplicación a creencias morales? La misma pregunta surge respecto de la verdad en otras disciplinas, como las ciencias naturales y la matemática, cuyos miembros, si no están corrompidos por una mala filosofía, se conciben a sí mismos trabajando dentro de una tradición intelectual que cambia con el tiempo, y tratando de descubrir lo que las cosas son realmente. Pero aun si hubiera una respuesta generalmente aceptada a la pregunta en estas áreas, lo que no ocurre, ella no sería transferible. Debemos preguntarnos sobre el tipo de objetividad que podemos aspirar a lograr para la moral en sí misma: ¿qué estándar independiente de aquello a lo que estamos habituados está implícito en la idea acerca de que esto puede estar seriamente equivocado, sea con respecto a la propiedad privada, la guerra, las relaciones entre los sexos, el castigo de los criminales o lo que fuera?

Alasdair MacIntyre cree que, en la cultura moderna, el intento de encontrar estándares éticos y políticos independientes de concepciones específicas e históricamente contingentes ha sido un desastre, y en su obra posterior a *After Virtue* (1981), prosigue con su ataque a la Ilustración. "¿De qué?", pregunta,

> nos privó la Ilustración? La Ilustración nos hizo ciegos, en gran medida, para algo que ahora tenemos que recuperar, y que es, como argumentaré, una concepción de la investigación racional contenida en una tradición; una concepción según la cual los estándares de justificación racional emergen y forman parte de una historia en la cual se los reivindica por el modo como trascienden las limitaciones y proveen remedios para los defectos de sus predecesores, dentro de la historia de esa misma tradición.

Esta afirmación aparece al principio, con cierto aire de desafío, pero no está claro en forma inmediata qué es lo que significa. El problema es encontrar una interpretación de la concepción que supuestamente hemos perdido, respecto de la que los liberales posiluministas, atacados por MacIntyre, están ciegos.

Hay una interpretación perfectamente natural de su enunciado sobre la investigación racional en tanto contenida en una

tradición, así como de otras cosas que él dice sobre cómo funciona una tradición, que es compatible con las pretensiones más fuertes a perseguir principios universales por medio del ejercicio de la razón, a saber, que no se puede pensar sobre ningún tema complejo, excepto en el lenguaje y con métodos de justificación, crítica y argumento, que se desarrollen gradualmente en el tiempo, y no se pueden entender ni las preguntas ni las respuestas sin haber internalizado aquel trasfondo. La formulación de MacIntyre se aplica, en este sentido, a cada descubrimiento matemático construido sobre el que lo antecede, y se aplicaría seguramente a los resultados de cualquier teoría ética moderna, independientemente de cuán racionalistas o universales sean sus ambiciones. En este sentido, no hay ninguna forma de pensamiento que *no* cuente como una tradición.

De modo que, aun prestándole a la posición de MacIntyre cierta aceptación espuria, esa interpretación banal puede no ser correcta. Pero tampoco puede serlo ninguna otra alternativa obvia: la interpretación realista, según la cual los estándares no sólo emergen de una tradición histórica, sino que son tan internos a ella que la elección racional entre diferentes tradiciones carece de significado. Se descarta esta interpretación, porque MacIntyre rechaza firmemente el relativismo. De modo que, si los liberales modernos lo rechazan también, no puede ser eso lo que ellos han perdido.

Lo que MacIntyre intenta con su afirmación de que la racionalidad debe estar arraigada en una tradición tiene que ser otra cosa, algo que dependa de un sentido del término especial, restringido, según el cual ninguna forma de pensamiento moral o político cuenta como una tradición si sostiene que todos pueden tener una razón para aceptarlo. En opinión de MacIntyre, no hay razones universales, hay sólo razones para este o aquel grupo. Aunque una tradición pretenda ser verdadera y eventualmente admita su derrota frente al mayor éxito de una tradición rival que trata con los mismos problemas, ni siquiera intentará defenderse con argumentos del tipo de que cualquier individuo racional debería poder aceptarla: es, esencialmente, particular. "El progreso en la racionalidad se logra sólo desde un punto de vista." Por ello, la larga tradición intelectual de la teoría ética secular moderna no es una tradición en el sentido de MacIntyre, y el liberalismo puede considerarse una tradición sólo en su apariencia de moral pública oficial

de las sociedades ricas de fines del siglo XX. Las teorías morales antiguas y medievales, en cambio, sí emergen de tradiciones, por lo que al menos tienen alguna posibilidad de ser verdaderas.

MacIntyre se describe a sí mismo como un cristiano agustiniano, lo cual ha modeado su idea de cómo debería ser una moral digna, una idea que aplica también a las morales seculares. Lo que le atrae de las tradiciones, en este sentido, es que proporcionan a sus miembros algo más básico que argumentos y justificaciones para apoyar sus convicciones morales: ellas, realmente, crean personas cuya naturaleza es tal que ciertas cosas cuentan como razones, justificaciones y críticas de conducta, sin ser cuestionadas. Generan disposiciones de pensamiento y carácter, de las que el razonamiento de esa clase depende, por lo que generan una confianza en los resultados que la razón sola no puede otorgar.

En cambio, los sucesores de la Ilustración se han dedicado a una infructuosa búsqueda de principios "que serían innegables para todas las personas racionales... Dentro de esa clase de filosofía académica, que es la herencia de las filosofías de investigación iluministas, la investigación de la naturaleza de la justificación racional ha continuado con un refinamiento cada vez mayor y con un desacuerdo que no disminuye". En otras palabras, toda esta plática no nos conduce a ningún lado, y sería mejor tener, desde el principio, un acuerdo fiable, entre los miembros de un grupo más pequeño, con una formación compartida, antes que los interminables desacuerdos entre todos.

Este tipo de concepción no es necesariamente religiosa, pero se la podría adoptar en el espíritu del romanticismo secular, poniendo el acento en la prioridad de instintos y sentimientos compartidos. Pero MacIntyre no es un romántico en este sentido: él cree en la verdad, pero cree que sólo se la alcanza por medio de una vía local particular, cuando se demuestra que una de las muchas tradiciones rivales es superior a las otras para tratar los problemas de la vida, en tanto que estas se ven forzadas a conceder que aquella tuvo éxito mientras que sus propios estándares fracasaron, algo que puede ser internamente evidente a través de lo que él llama "crisis epistemológica".

Veo el atractivo de esta posición: ¿por qué no olvidar los principios universales, y concentrarse en desarrollar alguna forma

particular de vida colectiva desde adentro? Esta es la única manera en que se crea algo valioso. Pero, poniéndonos del otro lado del cerco, creo que es mucho más fácil para un cristiano agustiniano dar este consejo, que para un no creyente, aceptarlo. ¿Por qué deberíamos confiar en que una vía históricamente contingente, con sus pautas internas, nos conduzca a la verdad, a no ser que Dios nos esté guiando? Si no creemos en esto y, no obstante, no deseamos caer en el relativismo o escepticismo desesperanzado, lo mejor que podemos hacer es tratar de desarrollar la capacidad de criticar y evaluar tradiciones particulares, incluyendo la que nos formó, con métodos que no presupongan su validez. Esto no es lo mismo que barrer con todos los logros y empezar de cero con algunos pocos axiomas morales "que serían innegables para todas las personas racionales" como en la parodia de MacIntyre. No hay axiomas, sino preguntas que tienden a ser lo más comprensibles posible, y respuestas de sofisticación creciente, que generan una amplia gama de acuerdos y desacuerdos, seguidas a su vez por más preguntas. El objetivo es construir gradualmente un punto de vista que todas las personas razonables pueden compartir.

Esto, también, es tradición, aunque no en el sentido de MacIntyre, y perfecciona sus métodos, principios y formas de razonamiento a través del tiempo, aunque algunos filósofos modernos hayan pretendido más de lo que, para un no creyente, era razonable esperar. No voy a desechar, sin ninguna razón, lo que las tradiciones morales más locales pueden ofrecer, como la eliminación de desacuerdos arbitrarios en el trato de diferentes grupos de personas, una de las herramientas más poderosas del pensamiento moral posiluminista. La persistencia de desacuerdos entre los participantes de esta empresa es esperable; el crecimiento en nuestra comprensión de los temas y el acuerdo en soluciones parciales a los problemas de una sociedad decente me parecen mucho más impresionantes, pese a los horrores antiiluministas, de los que el mundo está lleno. Para poder ver esto se necesita prestar atención al contenido actual de las teorías éticas que le hacen perder la paciencia a MacIntyre.

He hablado de la tesis general de MacIntyre, pero la mayor parte del libro consiste en el análisis de cuatro ejemplos históricos, que están concebidos para apoyar la tesis, mostrando

cómo el razonamiento moral siempre depende de una tradición y no puede ser comprendido o evaluado si no es dentro de su contexto histórico. Los ejemplos son el aristotelismo, culminando en Santo Tomás de Aquino; la tradición agustiniana; la Ilustración escocesa, eventualmente subvertida por Hume, y el liberalismo moderno (despojado de sus pretensiones universales y reinterpretado como el código tribal de los abogados constitucionalistas norteamericanos y los de su raza).

Una larga historia nos lleva desde los héroes homéricos, a través del teatro trágico, Pericles, Tucídides, los sofistas y las consecuencias desastrosas de la Guerra del Peloponeso, hasta el intento de Platón de unificar las pretensiones de perfección social e individual. Todo esto nos es muy familiar, pero el interés aumenta con el claro énfasis que MacIntyre pone en la cercanía entre los intereses de Aristóteles y los de Platón, en la dimensión teológica de la teoría ética aristotélica, y en cómo su concepción sobre la relación entre ética y política es más exitosa que la de Platón.

> No hay ningún estándar externo a la *polis* por el que se la pueda evaluar racionalmente respecto de la justicia o algún otro bien. Comprender qué es la *polis*, cuál es el bien que es su función alcanzar y hasta qué punto la *polis* de cada uno ha alcanzado ese bien, todos estos requieren la pertenencia a una *polis*. Sin pertenencia... a uno le faltarán inevitablemente elementos esenciales de la educación en las virtudes y de la experiencia de la vida de las virtudes, que es necesaria para esa comprensión.

La *polis* nos ayuda a realizar nuestro fin en el orden teleológico del cosmos: "Ninguna racionalidad práctica fuera de la *polis* es la contrapartida aristotélica de *extra ecclesiam nulla salus*". Y en Santo Tomás esta concepción de la virtud y razón práctica se integra con la concepción de Agustín de la gracia divina. Esta, según indica MacIntyre en su conclusión, es la tradición que él desea seguir, aunque, naturalmente, no puede recomendársela firmemente a todos.

Whose Justice? Which Rationality? contiene una larga historia del trasfondo religioso y legal de la filosofía moral en Escocia antes de Hume, y una explicación de la reacción naturalista humeana en contra de ella. Los lectores que no se han puesto en contacto previamente con el pensamiento de Sir

James Dalrymple of Stair no sabrán por qué. El material tiene algún interés histórico, pero se enfoca más en el fracaso de Hume en conseguir un cargo de profesor que en sus propias ideas.

La exactitud en referir lo que otros dicen no ha sido nunca un punto fuerte de MacIntyre. Con el principio de que tiene una presa fácil, tiende a ser muy poco preciso cuando ataca: la representación de la teoría ética de Kant en *After Virtue* nos hace desconfiar de lo que él dice sobre un texto filosófico. El tratamiento de Hume aquí no es mucho mejor. MacIntyre ignora la flexibilidad teórica de la psicología moral de Hume, porque decidió que Hume era un defensor del orden social inglés de su época y que la política dominaba su teoría moral. (La anglofilia humeana irrita a MacIntyre, quien refiere con satisfacción evidente el juicio posterior de Walter Bagehot de que este tránsfuga escocés no dominaba el estilo idiomático inglés.)

Es verdad que los ejemplos de Hume sobre objetos de orgullo incluyen preeminentemente la riqueza, propiedad y posición social, y que su análisis de la justicia se centra en las reglas de propiedad y contrato. Es también verdad que negó que la igualdad social y económica fuera un elemento de justicia. Y se podría agregar todavía, a la acusación de MacIntyre, que la teoría humeana permite explicar criterios discriminatorios en la moral sexual y el sometimiento de las mujeres. Pero nada de esto muestra que su psicología presuponga los arreglos sociales y económicos de su época. Al concebirse como un estudiante empírico de moral, obviamente tenía que ofrecer una teoría que explicase las convicciones e instituciones morales que él podía observar, pero que también podrían explicar cómo las creencias morales cambian bajo diferentes condiciones. El marco analítico que Hume propone es mucho más flexible que la teoría lockeana de los derechos naturales, por ejemplo, ya que requiere que comprendamos la interacción entre circunstancias externas y psicología individual, y el papel crucial de la convención, al determinar la forma exacta de los derechos de propiedad y la jerarquía social en un lugar y momento dados.

Hume llamaba "artificiales" a las virtudes como la justicia, que están esencialmente ligadas a las convenciones y que alcanzan su objetivo sólo a través de la conformidad general y no individual: él inventó el tipo de análisis moral que condujo al

utilitarismo de las reglas. La presentación de MacIntyre de la teoría humeana de las virtudes artificiales y su psicología le debe muy poco al original; él sabe que debe expresar cierta fidelidad social, de modo que no tiene que observar demasiado cuidadosamente lo que dice, lo cual no es un error inusual entre aquellos que nos urgen a prestar más atención a la historia. El dice que Hume cree que su justificación de las reglas de propiedad tiene validez en todo tiempo y lugar, ignorando su explicación de las limitaciones que imponen las circunstancias —escasez extrema, abundancia extrema, benevolencia universal o egoísmo universal— que hacen que las reglas sean inaplicables. Y dice, como si fuera una objeción, que "es posible que la inestabilidad y fractura social parezcan surgir en ciertas circunstancias, precisamente por la aplicación de las reglas de propiedad", mientras que esta es, precisamente la clase de cosa que para Hume resultaría en un cambio de las reglas y convenciones que definen la propiedad. Hume podría habernos proporcionado una explicación muy plausible sobre el desarrollo de la moral social del estado de bienestar (y también de la liberación de las mujeres) en estas líneas.

No defiendo la teoría ética de Hume: me parece demasiado flexible y receptiva, antes que lo contrario, es decir, demasiado limitada a la clase de crítica que le permite a una moral psicológicamente comprensible, o a un sistema estable de derechos de propiedad). Pero el tratamiento despectivo de MacIntyre contrasta con la caridad que exhibe al discutir la concepción de Aristóteles sobre el carácter natural de la esclavitud y la inferioridad de las mujeres, donde dice, correctamente, que: "el error de Aristóteles... fue que no entendió cómo la dominación de cierto tipo es, de hecho, la causa de aquellas características del dominado que luego se invocan para justificar la dominación injustificada", y agrega que este error puede "extirparse del pensamiento de Aristóteles sin negar sus tesis centrales sobre el mejor tipo de *polis*".

Los ejemplos históricos de MacIntyre ni siquiera apoyan y, menos aún, establecen su tesis central. Tal vez esté en lo cierto en que el único camino hacia la verdad es el de una tradición históricamente particular, que evita la búsqueda de fundamentos universales de aceptabilidad racional; pero señalar que existieron tales tradiciones, y que la búsqueda de fundamentos universales no terminó con los desacuerdos morales, no nos da

una razón suficiente para creer en ello. Vistas con beneplácito, todas las críticas de MacIntyre sobre la dificultad de traducir el lenguaje moral entre culturas, y la importancia, para toda comprensión cultural, de mirar el trasfondo histórico y entrar en el lenguaje, no son incompatibles con un racionalismo ético kantiano, independientemente de que esta concepción sea verdadera o no. En mi intepretación, es la religión de MacIntyre la que domina su filosofía. El quiere producir un argumento que no descanse en premisas religiosas, que muestre que sólo es posible algo como la moral religiosa. Es imposible hacerlo. Pero, para él, la conclusión del argumento es evidente, por otras razones.

Estamos enfrentados con problemas difíciles de justicia social, que el libro de MacIntyre no nos ayuda a tratar, ya que no contiene ningún análisis moral sustantivo. Hacia el final, respondiendo a qué deberíamos hacer si él tuviera razón, dice, "eso depende de quién es usted y de cómo usted se comprende a sí mismo". En otras palabras, encuentre su identidad y, con suerte, tendrá una convicción. Pero esto es inútil para responder a los problemas de cómo vivir decentemente en un mundo en el que las concepciones tradicionales como las que hay conducen a conflictos tanto como conducen a la armonía, y en las cuales muchos de nosotros realmente no sabemos qué les debemos a los demás seres humanos, y creemos que quienes piensan que ya lo saben están equivocados. La continua búsqueda de una solución a estos problemas más profundamente fundamentada se genera no sólo por una exigencia intelectual, sino también moral. MacIntyre pretende liberarnos de una ceguera, pero en verdad nos está pidiendo que regresemos a aquella ceguera, respecto de la dificultad del pensamiento moral, de la que la teoría ética, en lo que fue uno de los mayores logros, pudo escapar.

22

Kolakowski: La modernidad y el diablo*

Este ensayo fue una reseña de Modernity on Endless Trial *(University of Chicago Press, 1990), de Leszek Kolakowski.*

Este libro reúne veinticuatro ensayos escritos entre 1973 y 1986, que expresan la concepción de Leszek Kolakowski sobre Dios, el hombre, la razón, la historia, la verdad moral y el pecado original, incitada por la observación de la dramática lucha entre la cristiandad, la Ilustración y el totalitarismo moderno. Se trata de una maravillosa compilación de temas; lamentablemente, la mayoría de las partes de *Modernity on Endless Trial* dan la impresión de haber sido lanzadas sin mucha inversión de tiempo o reflexión. Están salpicados con rápidas declaraciones, como la siguiente, del ensayo que le da el título al libro:

> [E]n el sentido normal de "racionalidad", no hay más fundamentos racionales para respetar la vida humana y los derechos personales humanos que los que existen para, por ejemplo, prohibir el consumo de langostinos entre los judíos, de la carne el Viernes Santo entre los cristianos o del vino entre los musulmanes. Son todos tabúes "irracionales". Y un sistema totalitario, que trata a la gente como partes intercambiables en la maquinaria estatal, para usarlos, descartarlos o destruirlos, según las necesidades del Estado, es, en un sentido, un triunfo de la racionalidad.

* Reimpreso con autorización de *The Times Literary Supplement*, 14-20 de diciembre de 1990.

Las caricaturas de la razón que dan lugar a la fe, no contribuyen en nada en nuestra comprensión de la posición cristiana de Kolakowski. Y él sabe más: en un ensayo llamado "¿Por qué necesitamos a Kant", reconoce con respeto el intento de fundamentar derechos humanos universales en el razonamiento, en vez de basarlos en tabúes o contingencias históricas o psicológicas. La "razón" que figura en sus atolondradas observaciones es una invención de un filósofo, adaptada de Hume, y mucho más alejada del pensamiento ordinario que el concepto kantiano.

El punto de vista distintivo de Kolakowski hubiera merecido una profundidad intelectual mayor. El cree que la sobrevaluación de la razón es peligrosa, porque su forma más clara es meramente un método de cálculo, útil para descubrir medios para fines, pero bastante inaplicable a la evaluación de los fines mismos. Tiene, sin duda, la esperanza kantiana de un ejercicio "trascendental" de la razón, tendiente a descubrir una base menos arbitraria para los requisitos morales que los tabúes o preferencias dados, pero, pese a que Kolakowski admite esta posibilidad y admira su ambición, no evalúa los intentos de realizarla: aparentemente, él cree que una base no racional y religiosa, para la aceptación de leyes naturales, nos proporciona una mejor protección del nihilismo.

En "The Illusion of Demythologization", sostiene que la fe de la Ilustración es todavía posible, tanto lógica como psicológicamente; que no hay incoherencia entre nuestro conocimiento histórico y la creencia de que Jesús era el hijo de Dios, ni ninguna incoherencia entre nuestro conocimiento científico y la creencia en la inmortalidad. La afirmación iluminista del igual valor de todos los seres humanos era, según Kolakowski, una herencia de la tradición cristiana contra la que la Ilustración se rebeló. Pero la sobrevaluación de la razón y el abandono de los tabúes, que marcó aquella rebelión, finalmente condujo al rechazo de los absolutos morales no defendibles por medio de la razón, con resultados familiares y terribles. Se le echa también la culpa a la ilusión de perfección humana, que sucedió a la pérdida de creencia en el mal inherente a la humanidad, con los consiguientes intentos catastróficos de imponer políticamente una hermandad universal.

"Politics and the Devil" ofrece una breve historia de la cultura occidental moderna, como una lucha por el alma humana.

Kolakowski supone que la separación de política, ciencia, arte, filosofía y tecnología de la autoridad de la Iglesia, que es el comienzo de la Modernidad, no fue simplemente la labor del Diablo, sino que puede concebirse como una nueva *felix culpa*, que liberó nuestra creatividad y también nuestra maldad. La Ilustración también contribuyó a la purificación de la cristiandad, al reducir la participación de la Iglesia en política, guerra y adquisición de riqueza. Pero el Diablo hizo uso de ello como una fuerza antirreligiosa, conduciendo primero al humanismo y finalmente al abandono de la verdad en política, a favor del poder. En un giro final, inventó una nueva idea de verdad, que se expresa en las pretensiones absolutas de Estados ideológicos que son caricaturas de teocracias, sin las ataduras de las inhibiciones religiosas.

Yo simpatizo con dos aspectos de la cosmovisión de Kolakowski. Primero, él insiste en la importancia de la moral no contingente, una moral que les asigna un valor irreductible e insustituible a los seres humanos, en forma de ley natural o derechos humanos. Este aspecto rechaza tanto la posición marxista de que los derechos son, o bien una hipocresía (en la sociedad burguesa), o bien superfluos (en la hermandad universal del comunismo), como el nervioso relativismo que duda de aplicar valores occidentales (incluso es un error llamarlos "occidentales") al juzgar cómo otras sociedades tratan a sus miembros. En segundo lugar, Kolakowski reconoce el mal inherente a la naturaleza humana, y rechaza el pensamiento utópico; admite que se usó la doctrina del pecado original como una excusa para no hacer nada por resolver problemas sociales graves y evitables, pero cree que la idea de la perfección humana ha tenido consecuencias peores. El se denomina a sí mismo un socialista-liberal-conservador, pero la sustancia política de su posición es la de un demócrata social conservador: defensor de la libertad individual, pero que evita la felicidad como fin político, valorando tanto la seguridad social como la personal para todos, y considerando al mercado útil, mas no sacrosanto. Estos valores no son inusuales, pero Kolakowski teme que están amenazados por lo que él percibe como una tendencia modernista, ya sea considerando todos los valores arbitrarios, o renunciando frente a los conflictos que surgen entre ellos.

Me parece que subestima los recursos de la reflexión moral moderna. Estoy de acuerdo con la pregunta de si los estándares

morales universales tienen alguna base de fundamental importancia, tanto teórica como política, pero toda respuesta religiosa necesita tanta defensa y explicación como toda respuesta secular. Es bastante oscuro cómo la fe puede garantizar confianza en una moral, sin ninguna razón independiente para aceptar su contenido. Quizá la fe puede justificar la abstinencia de langostinos o vino, pero no es suficiente para justificar la lapidación por adulterio y, si lo es, igual no es suficiente por sí misma para justificar un mínimo social para los pobres, o la igualdad moral y política de los sexos a menos, por supuesto, que la fe sea sensible al contenido de lo que nos exige que aceptemos.

La razón o, más bien, el razonamiento como método, trata de encontrar los principios más generales posibles para explicar casos particulares y eliminar distinciones arbitrarias e inexplicables, pero puede aplicarse sólo en relación con juicios sustantivos sobre lo que se está investigando. Su autoridad en la evaluación de sistemas morales plantea cuestiones profundas: depende tanto de la existencia de una forma de razón específicamente práctica —razonando sobre lo que debemos hacer— como de la afirmación de que la razón no es sólo otra práctica socialmente condicionada.

Un defensor del método kantiano debe sostener que es legítimo exigir las razones que justifican una práctica social contingente, de tal modo que no sea legítimo dar vuelta el asunto y poner a la razón misma en cuestión, apelando a dicha práctica. La asimetría surge porque cualquier afirmación de que la corrección de lo que uno hace es, automáticamente, una apelación a su justificación y, por lo tanto, está sujeta a la crítica racional. Todos los caminos conducen a la misma corte de apelaciones, una corte a la que supuestamente podemos acceder todos. La razón es universal, porque ningún desafío que se ha hecho a sus resultados puede evitar apelar a la razón, al final, sosteniendo, por ejemplo, que lo que se presentó como argumento es realmente una racionalización. Esto puede echar abajo nuestra confianza en el método o práctica original, sólo dándonos razones para creer en otra cosa, de modo que finalmente tenemos que pensar en los argumentos para decidirnos.

Es difícil entender cómo la fe puede ocupar un papel similar al de la corte de apelaciones para la creencia moral. No es en absoluto un método de investigación, y no parece ser el tipo de

cosa para la que, cuando se responde a todos los desafíos hechos a sus afirmaciones, la apelación es inevitable. Es por eso que parece, desde afuera, insuficiente por sí misma para apoyar principios morales que pueden ser desafiados y debe ser comparada con alternativas, buenas o malas.

Lo más llamativo del pensamiento de Kolakowski es una obsesión por el significado moral y filosófico de la historia, algo que le resulta natural en virtud de sus raíces en la tradición cristiana y la marxista, que él rechazó hace tiempo. Es una carencia del temperamento analítico ahistórico el que este tema tan importante le sea ajeno; la historia de la ciencia desempeña un papel significativo al pensar sobre conocimiento y metafísica, pero la historia de la humanidad es raramente, en esta tradición, el foco de las investigaciones sobre la verdad moral y el conocimiento moral.

La relación entre la teoría moral y la interpretación de sucesos históricos es compleja, pero indudablemente existe, ya que la moral pretende revelar posibilidades para el desarrollo humano en la dirección de mayor armonía y decencia, que la gente puede ser inducida a buscar, precisamente debido a ese mismo descubrimiento. Cualquier teoría moral que se nos presente con semejante esperanza debe enfrentar su propia versión del problema del mal: ¿cómo es la compleja historia de nuestra especie compatible con la hipótesis de que la mayoría de las personas son capaces de decencia moral, una presuposición de cualquier moral que pretende dirigirse a todos? ¿Cómo puede la supuesta autoridad de la moral reconciliarse con su frecuente impotencia?

Quizá puedan la doctrina del pecado original y la creencia en el diablo ayudar a comprender la extraña mezcla de tendencias opuestas que caracterizan la era moderna: el crecimiento de la democracia e igualdad socioeconómica, mayor tolerancia, la disminución de la crueldad en algunas sociedades, como opuestos al totalitarismo moderno con su escala industrial de asesinato y tortura, y el desarrollo de la incineración masiva, como una opción militar. Pero el pesimismo freudiano, el reconocimiento de la inmadurez de la raza y el poder de las pasiones infantiles poco comprendidas también pueden ser útiles.

Se sugiere a veces que los horrores del siglo XX y el poder contemporáneo del nacionalismo refutan los valores de la Ilustración. No es verdad. Pero un creyente en la moral universal,

humanista, sea kantiana o utilitarista, tiene *algo* que decir sobre estos sucesos, tanto explicativo como condenatorio. Necesita una concepción de los seres humanos que acomode tanto su capacidad para la moral, como su capacidad para lo monstruoso. Kolakowski insiste acertadamente en que la concepción de que los hombres son naturalmente buenos, pero están corrompidos por la sociedad, no es creíble. Kant ataca el problema en *La religión dentro de los límites de la mera razón* e intenta reconciliar la posición de que la moral es la expresión más completa de la libertad humana, con el hecho de que el mal verdadero requiere también libertad. Hay una necesidad de teorías seculares del mal, compatibles con afirmaciones universales acerca de una moral positiva, y la exposición a las preocupaciones como las de Kolakowski pueden llevar a filósofos no acostumbrados a ello a pensar sobre el tema. Kolakowski mismo, sin embargo, se lo toma con tranquilidad.

Índice temático

A Matter of Principle (Dworkin) 244
A priori, conocimiento 76, 77, 84, 106 n.1
A Theory of Justice (Rawls) 13, 154-71
Aborto 15, 216, 224, 242, 248
Abridor/destapador de botellas, *ergon* de 148
Abstracción, científica 210
Abuso sexual de niños 52-3, 62n.10
Acción, 121; compromisos del yo futuro 237; comparación de posibles resultados 180; limitaciones en la libertad 182; deliberación sobre 221; libertad individual de 173; intencional 92-3; y justificación 58, 213-4; motivación para la 201, 202-3, 204, 207-8; subestructura fisiológica de 88; estructura primitiva de 122; reflexión y deseo 225; y derechos 181; a corto plazo 189
Activa, representación 236
Actos de habla 12
Acuerdo 76, 85; ético versus lingüístico 171 n.1; sobre reglas gramaticales 83-4; y justicia 159; en el sentimiento moral 216; sobre el uso de términos 75. *Véase también* Desacuerdos.
Adaptación, entre el juicio y la letra de la ley 247-8
Adjudicación 245, 252
Adquisición: del lenguaje (*véase* Lenguaje: adquisición de); de la propiedad, 178
Adulterio, lapidación por 266
Afirmaciones: conocimiento (*véase* Conocimiento, afirmaciones); morales 198-9
Afirmaciones inverificables 70-1
Afirmaciones morales, universalidad 198-9
After Virtue (MacIntyre) 255
Agustín de Hipona 259
Agustiniano, cristianismo 257, 258, 259
Alarmas de humo, 184
Albritton, Rogers 12
Alcohol 230; abstención de 233, 238; prohibición a los musulmanes 263
Algoritmo 142 n. 9
"Algunas lecciones elementales de psicoanálisis" (Freud) 24
Alma 127, 146, 150, 153 n.2
Altruismo 237
Amenazas, respuestas a las 109
Amor, objetos de 47, 48
Amorales 205
Analogías: entre interpretación

Agradezco a Nicholas Humez por compilar este índice.

jurídica y literaria 252; con otras mentes 37 n.10 y 11 (*véase también* Empatía)
Anarquía 176, 187 n.2
Anarquía, Estado y utopía (Nozick), 173-86
Animales: trato ético a los 186; vistos en las nubes 41
Anscombe, Elizabeth 119
Antimeritocracia 158
Antiperfeccionismo, humano 158. *Véase también* Perfectibilidad
Antirreduccionismo 11. *Véase también* Reduccionismo
Apariencia del tiempo, subjetivo 115
Apoderados, post mortem 239
"Appeal to Tacit Knowledge in Psychological Explanation" (Fodor) 87-94
Aquino, Santo Tomás de 259
Arbitrariedad 186
Arbitro, imparcial, obediencia a 160
Argumento 256; actividad de, versus producto 17; relevante 254
Argumento de la "habitación china" 138
Argumento del lenguaje privado 64, 67, 70, 75 n.2
Aristóteles 145-53, 220, 259, 261
Aristotelismo 259
Aritmética 137-9
Armas nucleares 234
Armstrong, David M. 95-106
Arrow, Kenneth Joseph 16
Arte 42, 265; custodia sobre 252; y significado 49-51; y suerte moral 213
Asalto 176, 181
Asesinato 176, 179, 182, 267
Aspects of the Theory of Syntax (Chomsky) 13, 77-82
Astronomía 211
Atarse los cordones 88, 89, 90, 91, 93
Atención médica. *Véase* Salud pública
Atomos 187-8 n.3
Atributos, no físicos 96. *Véase también* Propiedades de entidades; Cualidades
Atrocidades, morales 194
Austin, J. L. 12
Autoconciencia 115; histórica 254
Autoconocimiento 59, 62 n.14
Autodesprecio 45
Autodirigido, cohete 98
Autoestima 172 n.5
Autómata, inconsciente 130
Autonomía 147, 159, 240
Autoridad: moral 212; vinculante temporariamente 239
Axiomas, morales 258
Bagehot, Walter 260
Baloncesto, profesional 184
Basmajian, J.V. 90
Beneficios/perjuicios, equilibrio 180
Bentham, Jeremy 19
Bien 154; concepciones diferentes del 163, 164; igualdad como 185-6; justicia dependiendo del 163-4; medida del 202; contra lo correcto 155; concepción esquelética del 167. *Véase también Eudaimonía*; Equidad; Justicia; Corrección; Virtudes
Bienes: primacía de la libertad 158; primarios 171 n.2, 161, 163, 209 n.9
Bienestar, general 183, 251
Bienestar humano 154. *Véase también Eudaimonía*
Biología, molecular 120
Biológico, naturalismo 135
Biorretroalimentación 62 n.15
Bolsas de aire, automóviles 231
Bonaparte, Marie 23
Bote salvavidas 231, 232
Bradley, F.H. 48

Brown versus Board of Education 249
Brujería 60-1, 62 n.14, 133-4
Brutalidad 223
Buena fortuna 147
Búsqueda e incautación 188 n.6
Calendario, biológico 52
Cantor, George 16
Capacidad: de adquisición de conocimiento 82; de adquirir otras capacidades 80; innata, versus conocimiento innato 79
Carácter: de la vida 222; moral 216
Carga de la prueba 254
Carnap, Rudolf 11, 19
Cartesiana Linguistics (Chomsky) 13
Cascos 241
Cascos de motociclista 241
Castigo 187n.1; cruel e inusual 250
Causa: análisis de actividades mentales 97-101, 102; y efecto, semejanza de 28; generalización sobre 56-7; juicios sobre 41; de percepciones 98-9; física, de la conducta 96
Cauterización, salvar la vida por 236, 241 n.1
Cavell, Marcia 61 n.8
Censura 242. *Véase también* Pornografía
Cerebro: y mente 52, 96, 112; estados de 109, 119, 128-9. *Véase también* Estados mentales
Chamberlain, Wilt 184
Chocolate, gusto de 113
Choice and Consequence (Schelling) 229-34
Chomsky, Noam 13, 16, 33-4, 76-86, 136, 141
Churchland, Patricia S. 142 n.5
Churchland, Paul M. 142n.5
Ciencia 219, 228, 265; neutralidad de la 210; metodología de la 213-4; cosmovisión de la 14, 16-7
Clarke, Thompson 14, 15
Clima moral 254
Clínica, observación: evidencia a partir de 42
Coagulación de la sangre 87
Cobardía 223
Código, lenguaje 244, 246, 247-8
Coerción: libertad de 178; por el Estado 177
"Coercion" (Nozick) 187 n.1
Cohen, Marshall 15
Cohete, autodirigido 98
Cólera 147
Color 99, 106 n.1; percepción e identificación del 69-74, 116
Comparación, de a pares 155
Compensación 180, 187 n.1, 207, 236
Complejos totales 187-8 n.3
Comprensión: estética 49-50, 51; psicoanalítica 56-7; autoevidente 76; sentido de la 71; a través de la identificación del yo con otros 37 n.10 (*véase también* Empatía); utilidad de la 45, 46
Compromisos: por el incompetente 236; consigo mismo 237, 240-1; con el contrato social 168
Comunidades, filosóficas 15-6, 20
Concebir, posibilidad de 104-5
Conceptos a priori 120-1
Conceptos morales, lógica de los 196
Conceptos naturales 121
Conceptos: versus concepciones 249; éticos 223
Conciencia 94 n.3, 110-6, 126, 215; atribución de, al inconsciente 37 n.11; y punto de vista de primera persona 131; artículo perdido de Freud sobre la 31; en la adquisición del lenguaje 78-80; estados mentales de la 103-4; procesos físicos de 25-6, 30-1; Searle sobre la 130-7; versus

autoconciencia 115 (*véase también* Autoconciencia); y sensaciones 37 n.14; subjetividad de la 30, 31
Condiciones de asertividad 141n. 2
Conducta angustiada 109-110
Conducta preventiva 109
Conducta sexual 53, 176, 260
Conducta: patrones objetivos de 48; principios de 211; inconscientemente motivada 58; preventiva y aflicción 109; causada por estados mentales 103; peligrosa o injuriosa 187 n.1; discriminatoria 98; noble 236; causas físicas de 97; sexual 176.
Conductismo 18, 97, 98, 109, 114, 129-30, 142 n.3
Confianza, ética 225
"Conflictos de valor" (Williams) 216
Conflictos de intereses 206
Confrontación, "opcional" versus "real" 216-7
Conocimiento: y respuesta estética 51; a priori 76, 77, 84, 106n.1; y conciencia 92; ético, versus confianza 224; fáctico y científico 212; de los objetos superiores 151; innato 76, 77-82, 84; carencia de fundamento para el 84; de otras mentes 28, 37n.10 y 11, 67; supresión de (*véase* Velo de ignorancia); tácito 88-92; inconsciente 80
Conocimiento, afirmaciones del: del lenguaje 74; inverificable 70-1
Consciousness Explained (Dennett) 112, 113-6
Consecuencialismo, moral 199-200, 205
Consentimiento 160
Conservación, ontológica 120
Conservadurismo, 194; contraste con el liberalismo 247-8
Contemplación 145, 152; facultad de 147
Content and Consciousness (Dennett) 107-111
Conteo de las células blancas de la sangre 89
Contractualismo 219-20, 225. *Véase también* Contrato social
Contratos 260; violación de 176; restricciones sobre 182; derecho a firmar 237; social (*véase* Contrato social)
Contrato social 154, 156, 158-9; y elección 160-2; y justicia 158-9
Convencionalismo 245
Cooper, John M. 145
Corrección: como equidad 159; versus el bien 155; y justificación 266
Corredor, larga distancia 233
Corte Suprema, E.U.A. 242, 251
Creatividad 265
Credibilidad, intuitiva 58; moral 157
Creencias 29, 101, 108, 130, 140; falsas 99; justificación de 221; adquisición del lenguaje y 79; y capacidad de aprendizaje del lenguaje 83-6; moral 255, 266; en la objetividad 227-8; inconscientes 81
Creencias morales, verdad de las 256-7
"Crime, Punishment, and 'Pale Criminality' " (Wollheim) 48
Crimen 236; organizado 230
Cristianismo 264, 267; agustiniano 257, 258, 259
Crítica 256; racional 266
Crítico, pensamiento 190
Crueldad 267
Cualidades, Freud, sobre las 25. *Véase también* Atributos; Propiedades de entidades
Cuerpo: imagen del propio 118,

123; operaciones involuntarias del 87, 88; mente separada del 97, 103
Culpa 215
Curación 87
Dalrymple, James 260
Daños 179, 181; respuestas a los 109
Dar, voluntariamente 183
Darwin, Charles 16
Davidson, Donald 17, 57, 119, 122
De Anima (Aristóteles) 146
De a pares, comparación, de Rawls 155
De razón insuficiente, principio de 166
Debido proceso 250
Decisión 224
Deliberación: para la acción 221; moral 205-7, 215
Delirios, paranoicos 57
Dennett, Daniel C. 107-11, 112, 113-6, 142 n.2
Depravación 151
Depresión 45
Derecho natural 264, 265
Derecho: absoluto versus calificado 184-5, 188 n.10; a la propiedad 178
Derecho: como ayuda para el autocontrol 233-4; penal 48; de divorcio 238; aplicación, justificación del 246 (*véase también* Poder público); e identidad 235; como integridad 245, 248, 252; naturaleza interpretativa del 244; letra de la ley 244, 246, 248; natural 264; y autodominio 240-1. *Véase también* Jurisprudencia
Derecho penal 48
Derechos 175, 177, 205, 263; de criminales acusados 242; de gobiernos 177; jerarquía de los 178-9; visión libertaria 180-1; posición marxista 265; moralmente espurios 249; naturales 260, 265; de posesión 178, 240; privacidad 248; respeto por los 184; universalidad de los 263-4; violaciones a los 182, 188n.6
Derrida, Jacques 18
Desacuerdos, morales 197
Desarrollo: biológico, de la especie 52; infantil 48; psicosexual 58
Descartes, René 34, 39, 65, 104, 105, 127
Descubrimiento 228; afirmaciones de, experimento defectuoso 39
Deseos 29, 101, 122, 149, 150, 221, 225; para acciones morales 201, 202-3
Desesperación 97
Desigualdades de poder 174, 188 n.5, 250-1, 260, 266
"Desire, Belief, and Professor Grünbaum's Freud" (Wollheim) 43-4
Detención, preventiva 188 n. 6
Deuda, cancelación de 238
Diablo, el 265, 267
Dietas 233, 241
Digestión: almeja 148-9; humana 153
Dios 146, 258, 263
Dioses, *erga* de 149
Discriminación: arbitraria 258; inversa 242, 243, 247
Discriminación, escuelas 249
Disgusto 97
Distinción subjetivo/objetivo 135, 136
Distributiva, justicia 184-5, 186, 188 n.10, 232, 235, 236
Doble hélice 27
"Dogma del sentido de largo alcance" (Pears) 69
Dolor 66, 67, 99, 108, 109, 130, 179
Dolor de cabeza versus golpe en la cabeza 124

Dotes, naturales 185. *Véase también* Talento
Drama edípico, psicoanalítico 47
Dualismo: antipatía hacia 136; temor al 133; y materialismo 132-3; mente/cuerpo 104, 107, 109, 111, 114-5; propiedad 125, 127, 142 n.6; sustancia 127
Duda 66, 152. *Véase también* Escepticismo
Dworkin, Ronald 15, 18, 242-53
Educación 175
Eficiencia 185
Einstein, Albert 16,54
El concepto del derecho (Hart) 13
El imperio de la justicia (Dworkin) 242-53
El libro azul (Wittgenstein) 63-4
El libro marrón (Wittgenstein) 63-4
Elección 163, 165, 170: individual 156, 173-4; punto de vista en la 213; en el contrato social 156, 160-2
Eliminativo, materialismo 14, 133, 142 n.5
Emotivismo 73
Empatía 215-6
Empírica, realidad 136
Encarcelamiento 182
Enmienda 14ª 247
Entrenamiento de gatos 91
Envidia 172 n.5
Equidad 154, 155, 159, 170, 171 n.2, 250
Equilibrio beneficios/perjuicios 180
Equilibrio, reflexivo 157
Ergon 146-9; de perros y bienes 149; humano 151, 153
Erótica, vida 44, 55. *Véase también* Amor; Conducta sexual
Escepticismo 73-4, 188 n.5, 258; sobre teoría ética 227-8; sobre la separación de poderes 251; Wittgenstein, 67
Esclavos y esclavitud 60, 62 n.14, 152, 217, 237, 240, 261
Escocesa, Ilustración 259
Escuchas telefónicas 188 n.6
Escuelas: rezos en las 242; discriminadas 249
Especiecismo 224
Espectros, invertidos 99
Espermatogénesis 87
Esposa versus extraña, rescate 214
Esquema del psicoanálisis (Freud) 24
Establecimiento de una religión 250
Estado de bienestar 261
Estados mentales 119, 141 n.2; como causa de conducta 102-3; conscientes (*véase* Conciencia); experiencia de los propios 103; indescriptibles en el mundo de la física 133; internos 97; intencionalidad de 137; necesidad versus contingencia en los 106 n.1; concomitantes físicos de 26, 16 n.7; y estímulos 96; inconscientes 24-6, 140 (*véase también* Inconsciente/inconsciencia
Estilos de vida, individuales 176
Estímulos, y estados mentales 96
Ethics and the Limits of Philosophy (Williams) 218
"Ethics, Law, and the Exercise of Self-Command" (Schelling) 235-41
Ética a Eudemo (Aristóteles) 145-6
Ética a Nicómaco (Aristóteles) 145, 146, 151
Ética: aristotélica 222, 259; de asociaciones 250; emotivista 73; fundamentación de la 195; e historia 267; e identidad 235; intuicionista 156; kantiana 231, 260, 262; lingüística como mo-

delo para la 171 n.1; objetividad de la 210; en el estado de naturaleza 177; teorías de la 196, 210, 215-6, 218, 225, 228, 257; conceptos "gruesos" de la 223. *Véase también* Moral
Eudaimonía 145-53
Eutanasia 234
Eventos mentales 106 n.3; y eventos neurales 107
Evidencia 56, 221; a partir de la observación clínica 42; forense, evaluación por jurado 246; del significado 74; psicoanalítica 57
Evolución 109
Examined Life (Nozick) 176
Excelencia 151-2
Experiencia: y comprensión estética 51; consciente 112-3; de los estados mentales propios 103; perceptiva 126; y lenguaje público 70-1; del hablante, del lenguaje 82
Experimentos, defectuosos, afirmaciones de conocimiento basadas en 39
False Prison (Pears) 63, 64-75
Fama 17-8
Fanáticos 198
Fantasmas 133
Fe: en la igualdad 264; y creencia moral 266, 267
Feigl, Herbert 11
Feyerabend, Paul 18, 142 n.5
Filosofía 265; apetito de, para justificaciones 225; cultura de las décadas de 1960-1990 11-20; actitud de Freud hacia la 36n.1; de la matemática 75; de la mente 14; naturalidad de los problemas en la 14; sobrevaluación del lenguaje en la 189; popularizada 17-8; escritos que reflejan metodología 75 n.1
Física 120, 127; y descripción de estados mentales 133; Freud, sobre la 25; newtoniana, versión mentalista de la 34-5; de partículas 226; visualizaciones de la 25
Fisiología, y acción intencional 87-8
Fliess, Wilhelm 23-4, 36 n.1, 52
Fodor, Jerry 87-94
Foot, Philippa 13
Formas de vida, wittgensteinianas 64-5
Foundations of Psychoanalysis (Grünbaum) 43-4
Fraude 176
Frege, Gottlob 13, 19
Freud and his critics (Robinson) 52-6
Freud, Sigmund 123; correspondencia con Fliess 52; sobre homosexualidad y paranoia 57; como materialista 23-26; sobre la melancolía 45, 61 n.6; metodología de 39-43, 45, 46, 56, 58; sobre la filosofía 36 n.1; sobre los concomitantes físicos de la psicología 36 n.7; teoría del superyó 47, 48; éxito terapéutico de 53
Fried, Charles 15, 214
Friedrich, Caspar David 50
Fuerza, física 179. *Véase también* Asalto; Poder público
Fumar (tabaco) 233-8
Funcionalismo, mental 129, 142 n.3
Fundacionismo, moral, de MacIntyre 256-7
Futuro, compromiso con acciones en el 236
Gasolina, racionamiento de 234
Gatos, entrenamiento de 91
Gaugin, Paul 213
Generosidad 183
Genética y genes 96, 101
Gibson, Mary 171 n.3
Gobierno: compromisos del régimen anterior 238; subsidios

para la filosofía 18-9; mínimo, del estado de naturaleza 140 n.2; poder del 173, 176-7
Gödel, Kurt 16
Goffman, Erving 230
Goldwater, Barry 183
"Good Self and the Bad Self" (Wollheim) 48
Gorbachov, Mijaíl 44
Gramática 76, 138, 139, 156, 171; generativa 77-82; reglas de 80
Gratitud 223; por instituciones justas 171 n.2
Grice, Paul 12, 17
Grossman, William I. 35
Grünbaum, Adolf 39, 43-4, 46-7, 52, 53, 55, 56, 47-60, 61 n.4
Gusto, del chocolate 113
Hambre 130
Hampshire, Stuart 119
Hardware (wetware) del cerebro 130
Hare, R. M. 189-208, 220
Harman, Gilbert 13, 15
Hart, H. L. A. 13
Hegel, G. W. F. 19
Heidegger, Martin 19
Heisenberg, Werner 16
Hempel, Carl Gustav 11
Heterofenomenología 114
Historia 263; significado de 267; y teoría moral 267
Historicismo 261
Hobbes, Thomas 39, 170
Homosexualidad, reprimida 57
Hopkins, James 57
Hostilidad 29
Humanidades, subjetivismo en 18
Humanismo 224
Hume, David 85, 259-61, 264
Husserl, Edmund 19
"IA fuerte" 130
Identidad 95; y convicción 262; en derecho y ética 235; personal 236; teorías de la 102, 107, 129, 142 n.3. *Véase también* Yo
Identificación 29; imaginativa, de Hare 206; del yo con otros, 37 n.10 y 11, 40
Ignorancia 193; igualdad de 159-60, 162, 163, 167, 169
Igual distribución 184
"Igual protección", cláusula 247, 249
Igualdad económica. *Véase* Igualdad: de distribución
Igualdad: de distribución 184-6; económica 174, 260; confianza en la 264; de oportunidades 158; de influencia política 174, 188 n.5, 250, 266
"Illusion of Demythologization" (Kolakowski) 264
Ilusiones paranoicas 57
Ilustración 255, 263, 264, 267; escocesa 259
Imagen, mental 108, 111
Imaginación 40; fantasiosa 41-2
Imaginativa, identificación, de Hare 206
Imparcialidad 154, 198-200, 222; moral 194: radical 204
Imperativos categóricos 171 n.2, 201, 209 n.5
Imperativos hipotéticos 201, 208
Impuestos 183, 185, 188 n.7, 190-1; deducciones 234; progresivos 174; para reducción de riesgos 230-3
Impuestos progresivos 174
Incoherencia, en interpretación jurídica 249
"Inconsciente" (Freud) 24
Inconsciente/inconsciencia 25-6, 26; atribuciones de conciencia a 37 n.11; de los niños 40; conocimiento en el 80-1; estados mentales 24-6; filosófico 123 (*véase también* Intuición); imágenes mentales del 27-9. *Véase también* Conciencia

Incorregibilidad 97, 106 n.2
Indeterminación: principio de Heisenberg 64; en deliberaciones morales 215
Individualismo, liberal 164
Individuals (Strawson) 13
Indulgencia con los hijos 240
Infanticidio 216, 224
Inferencia 100
Insights: éticos 228; psicológicos 40, 43, 45
Instintos 257; de vida y de muerte 37 n.12
Instituciones: justicia de las 158-9, 161, 166-7, 171 n.2, 183, 188 n.10
Instrumento de adquisición del lenguaje de Chomsky 141
Integridad, la ley como 245, 248, 252
Intelecto 17
Intelectualismo 146, 150
Inteligencia artificial 87; "fuerte" 130
Inteligibilidad del lenguaje 66-7
"Intellectualism and Practical Reasoning in Aristotle's Moral Philosophy" (Cooper) 145
Intención 94 n.3, 221; abstracta versus concreta 249-50; y acción 92; del artista 51; y seguimiento de reglas 138-9; lingüística, y gramática 83-4; y la subestructura fisiológica de la acción 87-8.
Intencionalidad 141; de estados mentales 109-11, 137
Intereses, en conflicto, en deliberaciones morales 206
Interpretación 139; y arte 51; judicial 244; literaria y judicial 252
Introspección 100
Intuición 123; moral 156, 192; credibilidad de la 58
Intuicionismo 156-7, 189-90
Inversa, discriminación 242, 243, 247
Investigaciones filosóficas (Wittgenstein) 13, 33, 63, 123
Involuntaria, servidumbre 237. *Véase también* Esclavos y esclavitud
Irracionalidad, neurótica 60. *Véase también* Racionalidad; Razón
Jackson, Frank 142 n.4
Jirafa 149, 150
Jornales, contratos de 237
Jueces 244
Juego suma cero 234
Juicios: prácticos 226; en equilibrio reflexivo 157
Jurados, convicciones empíricas de los 246
Jurisprudencia 242-53. *Véase también* Derecho
Justicia: dependiendo de lo bueno 163; distributiva 232, 235, 236; como equidad 155, 159, 170, 171 n.2; de las instituciones 158-9, 161, 166-7, 171 n.2, 183, 188 n.10; objetividad de la 164; principios de 157, 164, 171n.2, 183-4; en propiedad y contrato 260; y contrato social 158-9; en la sociedad 262
Justificación 83-6, 85-6; de creencias 221; en casos legales controvertidos 244; y acuerdo intersubjetivo 76; para aplicación de la ley 246; de la moral 219; de neutralidad 163; de la propia conducta 48; apetito filosófico por la 225; retroactiva 213-4; de lo correcto 266; estándares de 255-6
Kant, Immanuel 63, 65, 171n.2; imperativos categóricos 171n.2, 201, 209 n.5, 212; sistema ético de 221-2, 228, 260, 262; sobre

libre albedrío y el mal 268; imperativos hipotéticos 201, 208; metodología de la razón 264, 266-7; sobre la percepción 26-7
Klein, Melanie 47, 48
Kolakowski, Leszek 263-8
Kripke, Saul 13, 15, 72, 73, 95, 106 n.1, 106 n.3, 120, 142 n.4
Kuhn, Thomas 18
La interpretación de los sueños (Freud) 24
La religión dentro de los límites de la mera razón (Kant) 268
"La verdad en el relativismo" (Williams) 216
Language and Mind (Chomsky) 34
Lapidación por adulterio 266
Lapsus de memoria 41
Lapsus lingüísticos 41, 55
Large Enclosure near Dresden (Friedrich) 50
Latidos del corazón 62 n.15, 89
Lazos familiares 41-2
Legislativo 251; y voluntad de la mayoría 249
Legitimación, del poder del Estado 177
Legitimación, política 141, 250. *Véase también* Contrato social
Lenguaje figurativo 37 n.12
Lenguaje: adquisición del 77-82; teoría del lenguaje de Austin 12; y pensamiento complejo 256; figurativo 37 n.12; bases para la corrección en el uso del 70-1, 72-3, 74; capacidad innata de aprendizaje, 76; legal (*véase* Derecho: letra de la ley); mental, oraciones del 108; mal uso del, en juicios éticos 192; moral 192, 197, 262; natural 33-4; sobrevaluación del, en filosofía 189; filosofía del 55; privado 64, 67, 70-1, 75 n.2; semántica versus sintaxis 138; lapsus lingüísticos 41, 55; experiencia consciente del hablante del 82; técnico, términos comunes 121; Wittgenstein 33, 66-7, 69-72
Lenguajes naturales 33-4, 74, 78
Leyes de divorcio 238
Leyes sobre dietas, religiosas 263
Liberalismo 154, 173, 225, 242, 243, 248, 255, 259; contraste con el conservadurismo 247-8; igualitario 158, 169
Libertad 158, 161, 167, 172 n.5, 173, 185, 240; de acción, restricciones sobre la 182; privación de la 179; igual 169; respeto por la 159, 183-4, de otros; de prensa 248-9, 250; racional 221; de la voluntad 122-3
Libertarismo 173, 173-86, 191
Licencias, morales 202, 204
Lingüística 33, 77-86; como modelo para la ética 156, 171n.1; teoría de la, y estructura física 24
Locke, John 188 n.4, 260
Locomoción 150
Lógica, habilidad para la 76
Lógico, positivismo 18
Lotería 160
Lucero del alba 104
Lucero vespertino 104
MacIntyre, Alisdair 38n.19, 254-62
Madre Teresa 195
Madre y superyó 47-8
Mal: problema del 267; teorías del 268. *Véase también* Maldad
Malcolm, Norman 12
Maldad 265 *Véase también* Mal
Malicia, de la tostadora 41
Marx, Karl 186
Marxismo 265, 267
Más allá del principio de placer (Freud) 37 n.12
Mason, Jeffrey 52-3
Matar 179; matar al padre, deseos de 29. *Véase también* Asesinato

Matemática 16-7, 220, 228, 256
Materia, conectada con la mente 117
Materialismo 120; de Armstrong 100-6; el modelo "cerebro como ordenador" 125; cartesiano 115; de estado central 97, 106 n.1; búsqueda compulsiva de teorías 132-3; y dualismo 133; eliminativo 14, 95, 133, 142 n.5; de Freud 23-36, 40; reduccionista 128, 130.
Materialist Theory of the Mind (Armstrong) 95-106
Maxwell, James Clerk 16
McDowell, John 223
McGinn, Colin 72, 106 n.7
Mecánica newtoniana, versión mentalista 34
Mecanismos de defensa 47
Melancolía 45
Memoria, lapsus de 41
Mental, imagen 108, 111
Mente/cuerpo, dualismo 104
Mente: y biorretroalimentación, 62 n.15; y cerebros 52, 96, 112; simulaciones por ordenador de la 87, 93-4, 94 n.4; conexión con la materia 117; incorpórea 97, 103; y conceptos naturales 121; otras 28, 37n.10 y 11, 67, 104-5; solipsista 67; estado de la, hacia la ética 224; teorías de la 23, 75, 105. *Véase también* Identidad; Estados mentales; Yo
Mentir 194, 223
Mercado 265; libre 188 n.4
Mercado negro, monopolios 230-1
Metodología: psicoanalítica 39-44, 45, 46, 55, 58; de la razón, Kant 264, 266; reflejada en escritos filosóficos 75 n.1; científica 114
Michelman, Frank 15
Mill, John Stuart 170
Mind and Its Depths (Wollheim) 39, 40-50
Mínimo social 167, 174
Minorías, protección de las 251
Mitigar el hambre 203
Moby Dick (película, basada en la obra de Melville) 189 n.1
Modernidad 265
Modernity on Endless Trial (Kolakowski) 263-8
Molécula de ADN 27, 96, 101
Molecular, biología 120
Moléculas 188 n.3
Monismo 133
Monopolios, mercado negro 230-1
Monstruosidad, capacidad para la 268
Moral, carácter 216
"Moral Complications and Moral Structures" (Nozick) 187 n.1
Moral, credibilidad 157
Moral, lenguaje 192, 262; lógica del 207
Morales, axiomas 258
Morales, juicios 199, 223; mal uso del lenguaje en 192; y razón práctica 201Morosidad 233
Moral Luck (Williams) 210-7
Moral Thinking (Hare) 189-95, 197
Moral, sentimiento, compartido 216
Moral: 42, 47-8, 49, 61 n.8; capacidad 268; contenido 202, 212; fundamentos 189; justificación 219; legal 243; tradiciones locales 258; no contingente 265; ordinaria 194; política 243, 245; sexual 260 (*véase también* Conducta sexual). *Véase también* Etica
Motivación 193, 201-2, 220; en el conjunto 229; conflictos intrapersonales de la 235-41, 229; modificada por la razón 215; proyectos de vida versus altruismo 214; para la acción moral 204-5; teoría de la 207

"Mourning and Melancholia" (Freud) 61 n.6
Movimiento por los derechos civiles 15
Muerte: instinto 37 n.12; probabilidad de 231
Mujeres: la posición de Aristóteles sobre las 261; sometimiento y liberación de las 260-1
Museos, donaciones a 239
Nacimiento, como contingencia de oportunidad 158
Nacionalismo 267
Narcóticos 230. *Véase también* Consumo de drogas
Naturaleza humana 153; compuesta 145; teleología de la 222-3; universalidad de la 50. *Véase también Ergon*: humano
Naturaleza: objetiva 35; estado de 177, 187 n.2
Naturalismo: biológico 135; moral 48
Necesidad 95, 106 n.1 y 3
Necesidad, mayor, y trato preferencial 206-7
Negociaciones, intrapersonales, aplicadas al grupo 236-7
Neurofisiología 96, 102
Neuronas 136; tercera clase, Freud 30, 37 n.14; motrices 90
Neuronas omega (ω), de Freud 30, 37 n.14
Neurosis 61 n.4; compulsiva 132; síntomas producidos por 41, 59, 60, 62n.14
Neurosis compulsiva 109-10, 132
Neutralidad y justificación 163
"Newcomb's Problem and Two Principles of Choice" (Nozick) 187 n.1
Newton, Isaac 16
Nietzsche, Friedrich Wilhelm 19
Nino, Carlos 16
Niños 152; indulgencia con los 240; compromisos de 236; favorecer a los propios frente a extraños 202-3; abuso de 52-3, 62 n.10; en la teoría del desarrollo de Klein 47-8; pasiones de los 267; sensaciones de los 116; y el inconsciente 40
Nous 146, 157 n.2. *Véase también* Mente
Novocaína, ilegalidad de la 230
Nozick, Robert 15, 19, 173-86
Nubes, animales vistos en las 41
Nutrición 150
Objetividad 56, 114, 116, 134, 164, 211; creencia en la 228; de la ética 210; de la justicia 164; kantiana 228; científica 228; científica versus razón práctica 219; en patrones de conducta 48; y términos universales 74
Objetos: de orgullo 260; de la razón 150
Observación, clínica, evidencia a partir de la 42
Observador, imparcial benévolo 157
Observador/distinción cosa observada 139
Olfato, sentido del 52
Ontico, bulto 107
Ontológica, conservación 120
Operaciones, corporales 93-4; elementales 89; involuntarias 87, 88
Oportunidades, igualdad de 158
Oraciones, en lenguaje mental 108
Ordenadores 137-8; cerebros como 125; relación hardware/software 113; estados internos de los 109, carencia de mente de 126; simulación de mentes por 87, 93-4, 94 n.4 (*véase también* Inteligencia artificial); Von Neumann 115
Ordenes sociales pluralistas 171
Organos, vitales 188 n.5

Orgullo, objetos de 260
"Oro", como concepto 121
O'Shaughnessy, Brian 117-24
Otras personas: interpretadas como otros yoes 37 n.10 y 11, 40; conocimiento de 28, 67; irracionalidad neurótica de 61
"Otro tiempo, otro lugar, otra persona" (Williams) 217
Ovulación 87
Owen, G.E.L. 13
Padre como superyó 47, 54
Painting as an Art (Wollheim) 40, 50-1
Palabras 70; corrección en la aplicación de las 70, 72-3, 74-5; términos no referenciales 108
Parcialidad 214; límites 215; de gente común 194, 195
Parfit, Derek 15, 204-5, 234, 236
Partículas, física de 226
Pasiones 150; infantiles 267
Pasiva, representación 236
Patas de cerdo, abstención de 238
Paternalismo 240-1
Patrones de conducta 48
Pears, David 13, 63, 64-75
Pecado original 263, 265, 267
Peirce, C. S. 19
Pensamiento 108; complejo, lenguaje requerido por el 256; crítico 190; moral 193; tradiciones del 256
Pensamiento moral, crítico versus intuitivo 193
Percepción 101, 149; del color 116; como experiencia 126; Freud, sobre la 25; e imágenes mentales 111; y propósito 97-9
Percepción profunda 88
Pérdida del objeto 45
Perfectibilidad, humana 264, 265
Perros, *erga* de 149
Persona: individual, acción adscripta a 93; inviolabilidad de la 178; común, parcialidad de la 194; otras (*véase* Otras personas); pobre 231; temporalmente extendida 234; total, soberanía de 235-41
Personalidad 93-4; presunción de 105; teoría hilomórfica de la 153 n.2
Personas trastornadas, compromisos hechos por 236
"Persons, Character, and Morality" (Williams) 214
Pesimismo, freudiano 267
Philosophy & Public Affairs 15
Pietismo 50
Pisa, torre inclinada de 140
Planes de vida 169, 214; ignorancia de los 164-5; perspectivas de, influencia de instituciones 166
Platón 259
Pobres 231
Pobreza 183
Poder: del gobierno 173, 176-7; desigualdades en el 174, 188n.5, 251, 260, 266; público, coercitivo 154, 177, 244-5
Polémica 18
Polis 259, 261
Política 42, 265; y moral 243, 245
Political Liberalismo (Rawls) 154
"Politics and the Devil" (Kolakowski) 264
"¿Por qué necesitamos a Kant?" (Kolakowski) 264
Pornografía 53, 243. *Véase también* Censura
Posibilidad 106 n.3
Positivismo 18; incoherencia en el 217; versus subjetivismo 55
Postura, erecta 52
Pragmatismo 245
Preferencias: de otros 205; satisfacción de 199
Prejuicios: raciales 247; subjetivos 134

Prensa, libertad de 248-9, 250
Prescriptivismo universal 191, 192, 200-5
Prescriptivismo, universal 191, 192, 200-5
Presión sanguínea 62 n.15
Prevención, coste de la 231
Preventiva, detención 188 n.6
Primarios, bienes 209 n.9
Primera Enmienda 248
Primera persona: deliberación fáctica no limitada a la 221; plural, de Wittgenstein 217; punto de vista 131; reflexiva, como agente de deseos 226
Principio de conexión, de Searle 140
Principio de diferencia 166, 172 n. 5, 168-9, 172 n.7
Principio de indeterminación de Heisenberg 64
Principio de resultado final de la justicia distributiva183-4, 188 n.10
Principio histórico de justicia distributiva 183
Principio maximizador, en deliberaciones morales 206-7
Principios: impersonalidad de los 214; legales 246; objetivos, en ética 220; verdad de los 228
Principios igualitarios 167-8, 172 n.7, 236, 264; y liberalismo 158, 169
Privacidad 173, 248
Probabilidades 166-7; de muerte 231
Problema mente/cuerpo 28, 126-8; Dennett, sobre el 107-111; Descartes, sobre el 103; O'Shaughnessy, sobre el 118-23. *Véase también* Estados mentales: concomitantes físicos de los
Procesos involuntarios, corporales 87, 88
Procesos mentales, concomitantes físicos 41
Procreación 148. *Véase también* Reproducción
Profunda, psicología 37 n.42
Prohibiciones, legales 177, 187n.1, 230
Proliferación, nuclear 234
Promesas a sí mismo 237
Propiedad 260; restricciones en el uso de 182; derechos de 178, 190; amenazas a la 179
Propiedad, dualismo de 125, 127, 142 n.6
Propiedades de entidades 128. *Véase también* Atributos; Cualidades
Propósito, y percepción 97-9
Propósito: *Véase también* Intención
Prostitución 230
Proyección, psicológica 51
"Proyecto para una psicología científica" (Freud) 23-4, 29, 37n.14
Prudencia 156, 204; versus regla maximin 165
Prueba, carga de la 254
Psicoanálisis, afirmaciones científicas del 40-3
Psicología: del sentido común 40, 41, 46, 57-8, 100; profunda, 37 n.12; folk 133; freudiana 23-61
Psicologismo 33
Psicopatología, cotidiana 46
Pública, escuela: rezos 242; discriminadas 249
Punto de vista de tercera persona 114
Pureza 14
Puritanismo, posterior a los años ochenta 53
Putnam, Hilary 142 n.3
Química 128
Quine, W. V. 13
Racionalidad 147, 255-6; y derechos 263. *Véase también* Razón

Racionalismo 78
Racionamiento, gasolina 234
Rawls, John 12, 13, 14, 17, 19, 154-71, 209 n.9, 220
Raza 247, 249
Razón 149, 150, 263; y cambio como motivación 215; libertad de la 221; posición kantiana 264, 266; sobrevaluación de la 264; práctica 201, 204, 219, 228; al servicio de las funciones inferiores, 150. *Véase también* Racionalidad
"Razones internas y externas" (Williams) 215
Razones, universales 256
Realidad 219; empírica 134, 136; moral 211; incognoscible, según Freud 25
Realidades morales, externas 211
Realismo 11, 211, 228
Reasons and Persons (Parfit) 234
Reconocimiento, del conocimiento inconsciente 81
Reduccionismo 120, 128, 141 n.2; materialista 130; mente/cuerpo 113; realista versus conductista 98
Refuerzo, de gatos 91
Regateo 230
Regla maximin 165, 167
Reglas: en conocimiento expresable 79-81; seguimiento de, como proceso intencional 138-9; gramaticales, 82, 83-4; Wittgenstein sobre 73, 74-5
Rehnquist, William H. 248
Reichenbach, Hans 11
Relativismo 216, 256, 258, 265; fácil, atribuido a Wittgenstein 64; calificado 223
Religión 42; cláusula de establecimiento de 250; y ley natural 264; supersticiones de la 60, 62 n.14; supresión de minorías 251
Renta, control 234
Representación, activa versus pasiva 236-41
Represión, psicoanalítica 53, 58, 61 n.4, 81
Reproducción 150. *Véase también* Procreación
Respuesta estética 50-1
Restricciones, implícitas en el liberalismo 154
Resultados de acción, comparativos 180, 181
Retroalimentación, control: del propio cuerpo 62 n.15, de cohetes 98
Revoluciones, intelectuales 39
Rezo en las escuelas 242
Riesgos 236; reducción de, a costa de impuestos 230-2
Riqueza, privada y hereditaria 174, 188 n.5
Robinson, Paul 52-6
Robo 176
Romanticismo, secular 257
Rorty, Richard 18, 142 n.5
Ryle, Gilbert 142 n.3
Sachs, David 61 n.4
Sacrificio 168, 172n.6, 179, 180
Salud 147
Salud pública 175, 231, 233
Sangre, coagulación de la 87
Sartre, Jean-Paul 19, 41
Satisfacción de preferencias 199
Scanlon, T. M. 15, 209 n.9, 220
Scheffler, Samuel 61 n.8
Schelling, Thomas 229-41
Schopenhauer, Arthur 19, 65
Schweitzer, Albert 195
Searle, John R. 15, 18, 125-41
Secuencias, simbólicas 137
Sed 129
Seguridad: personal 265; pública 231-2; social 241, 265
Selección natural 109
Semántica versus sintaxis 138

Semejanza de causa y efectos 28
Sensaciones 101, 104, 122, 149; de los niños 116
Sentido común: materialismo del 95; psicología del (*véase* Psicología: sentido común)
Sentido de largo alcance, dogma de 69
Sentido, de largo alcance, "dogma" de Pears 69
Sentimiento: moral 189-90; compartido 257
Ser omnisciente 193. *Véase también* Dios
Servidumbre involuntaria 237. *Véase también* Escalvos y esclavitud
Sidgwick, Henry 19, 192
Significado: y arte 49-51; condiciones empíricas para el 63; evidencia de 74; de la historia 267; psicológico 54-5; teoría del 137; de palabras 70
Símbolos, secuencias de 137
Simon, Bennett 35
Simplicidad en ética 195
Sintaxis versus semántica 138
Sistema de mercado libre 188 n.4
Sistema moral, sentido del 250
Smart, J. J. C. 98, 142 n.3, 211
Sobre la afasia (Freud) 36 n.7.
Socialismo, e impuestos 190, 191
Sociedad de Filosofía Etica y del Derecho (The Society for Ethical and Legal Philosophy) (SELF) 15
Sociedad: democrática, ideal de una 173-4; pluralista 171
Sociedad, voluntaria 159
Solipsismo 67
Sonido 36 n.9
Sophia 151
Sperry, R. W. 16
Spinoza, Benedict 142 n.6
Stich, Steven 142 n.5
Strategy of Conflict (Schelling) 229
Strawson, Peter 13, 133-4
Stroud, Barry 15
Subjetividad 68, 115, 131
Subjetividad ontológica 131
Subjetivismo, fácil 18, 55
Sueños 41
Suerte, moral 213-7
Suficiencia empírica 100
Sugestión, psicoanalítica 44, 59
Sujeto, conciencia del. *Véase también* Conciencia; Autoconciencia
Sulloway, Frank 52, 55
Suma 66
Superyó 47-8
Suposiciones, y capacidad de aprendizaje de la lengua 83-6
Sustancia, dualismo de 127
Swartz, Robert 13
Tabúes 263, 264
Tabúes de alimentación 238
Taking Rights Seriously (Dworkin) 243
Talento, como contingencia de oportunidad 158. *Véase también* Dotes
Tarski, Alfred 11
Tecnología 265
Teleología: aristotélica 228; de la naturaleza humana 222-3; y la *polis* 259
Tendencias, innatas versus conocimiento a priori 84
Tentación 193
Teoría de juegos 234
Teoría de la identidad neural 107
Teoría del conocimiento moral, Ilustración 219
Teoría del derecho 187 n.1
Teoría del valor del trabajo, de Marx 186
Teoría moral 156. *Véase también* Ética
Teoría: modelos *como si* 33 (*véase también* Analogías); físicas, de-

pendiendo de conceptos no fenoménicos 32; psicológicas, y estructura física 23
Términos no referenciales 108
Términos universales, corrección en su aplicación 70, 71-3, 74-5
Terroristas, armados con armas nucleares 234
The Language of Morals(Hare) 197
The Nature of Rationality (Nozick) 173
The Rediscovery of the Mind (Searle) 125-41
The Thread of Life (Wollheim) 40
Thomson, James 13
Thomson, Judith Jarvis 15
Tiempo 115
"Tigre" como concepto 121
Titanic, hundimiento del 232
Tolerancia 267
Toma de conciencia, darse cuenta 111; del propio cuerpo 117
Torre inclinada de Pisa 140
Tortura 182, 193, 267
Tostadora, maliciosa 41
Totalitarismo 263, 267
Tractatus Logico Philosophicus (Wittgenstein) 65, 217
Traición, personal 194
Transferencia de propiedad 178
Transferencia, psicoanalítica 44
Trastorno temporario, compromisos durante el 236
Trastornos 236, 240
Turing, Alan Mathison 16, 142n.10
Unanimidad de consentimiento 160
Unidad moral, persona total como 236
Uso de drogas 176. *Véase también* Narcóticos
Utilidad promedio, principio de 165
Utilitarism: For and Against (Smart y Williams) 211
Utilitarismo 155, 158, 168, 190-200, 203, 207, 211, 219; versus parcialidad natural 195; preferencia por el 199
Utilitarismo de la regla 261
Utilitarismo de preferencias 199
"Utilitarismo y autoindulgencia moral" (Williams) 216-7
Utopías, libertarias 186
Validez, y fuerte influencia cultural 59, 62 n.14
Valor: igual, de todos los seres humanos 264; de la vida 231; teoría del valor del trabajo de Marx 186; moral, cantidad de 200
Velo de ignorancia 159-60, 162, 163, 167, 169
Venus (planeta) 104
Verdad 56; moral 255, 263; de principios 228; simultánea, de teoría física y psicológica 23
Vergüenza 97
Viajes aéreos, seguridad en 231, 232, 234
Vida: custodia de la propia 240; instinto para la 37n.12; valor de la 231
Vietnam, guerra de 15
Violación de contrato 176
Virtud moral 152
Virtudes 222; justicia 250 (*véase también* Justicia); moral 152; supererogatorias 194
Vista 130
Visualización 32; en las ciencias 27-9
Vitales, órganos, trasplante forzado de 188 n.5
Volición 118. *Véase también* Voluntad
Voluntad 118, 119, 211; de la mayoría popular 245, 251; teoría de la 122
Voluntad de la mayoría 249
"Voz" 108

Walzer, Michael 15
Whose Justice? Which Rationality? (MacIntyre) 254-62
Wiggins, David 15
Will (O' Shaughnessy) 117-24
Williams, Bernard 15, 19, 210-28
Wittgenstein, Ludwig 12, 13, 19, 119, 123; sobre la primera persona del plural 217; sobre el lenguaje 33, 66-7, 69-72; sobre el significado 63; sobre las reglas 73, 74-5; sobre el escepticismo 67
"Wittgenstein y el idealismo" (Williams) 217
Wollheim, Richard 39, 40, 57; sobre el arte 50-1; sobre la moral 47-9; sobre la metodología psicoanalítica 43-4
Yo corpóreo 47, 123
Yo: múltiple 238; de los padres 241 (*véase también* Superyó); unidad del 233. *Véase también* Identidad

BEG
FILOSOFÍA

THOMAS NAGEL

Otras mentes

Acerca de este libro, Daniel Dennett escribió: «Su defensa sucinta de la razón como el ultimo y mejor árbitro es tan buena, que es necesario tenerla a mano como un extinguidor contra incendios (...): *La razón es universal porque ningún intento por cuestionar sus resultados puede dejar de apelar a la razón en última instancia*».

En estos análisis agudos y polémicos, uno de los pensadores más destacados del momento discute las teorías de Nozick, Dworkin, Dennett, Searle y muchos otros, especialmente en el contexto del intenso trabajo que Nagel ha emprendido sobre la definición del papel de la mente y la psicología en la ética.

Otras mentes reúne los ensayos críticos y reseñas más relevantes de Nagel sobre filosofía cognitiva, ética y filosofía política. Los trabajos analizan una serie de pensadores desde Aristóteles hasta Wittgenstein y algunos de los teóricos más significativos del campo de la filosofía política y del derecho, como Robert Nozick y Ronald Dworkin. Otros ensayos dan cuenta de la constante participación de Thomas Nagel en el debate en torno a la relación mente-cuerpo, entre los que encontramos respuestas muy sugerentes a Daniel C. Dennet, John Searle y otros. A lo largo de *Otras mentes*, Nagel expresa su convicción de que las cuestiones intelectuales que más nos afectan en nuestros días —desde la fundación científica de la teoría de Freud hasta las vicisitudes de la interpretación del derecho— son esencialmente problemas filosóficos.

Escritos en un estilo accesible, polémico y vivaz, estos textos representan lo mejor de uno de los pensadores más destacados del presente.

Thomas Nagel es uno de los filósofos contemporáneos más visibles, francos y comprometidos. El *Times Literary Supplement* lo calificó como «el filósofo más interesante de nuestros días». A lo largo de los últimos veinticinco años, Nagel ha sido una figura central en las polémicas públicas sobre subjetividad y conciencia, objetividad y ética, liberalismo y razón. Ha escrito en muchas publicaciones de carácter popular, académico y legal, participando así en un intenso intercambio de ideas y contribuyendo a poner en contacto las cuestiones contemporáneas de interés general con la tradición filosófica.

LECTURA Y CULTURA SIN EGOÍSMO
BIBLIOTECA
www.lecturasinegoismo.com

ISBN 978-84-9784-384-3
9 788497 843843
892010

gedisa editorial

CREATIVE COMMONS

PVP 9,90 €